

-299-

Identities = 50/58 (86%), Positives = 56/58 (96%)

Query: 137 LHKIELGCDYDNKQSQAVARKLGFTLEANTIRDRRDQGRKCDMRPGLLRSEWEKKER 194
 LHKIELGCDYDNKQSQAVARKLGFTLEAN RDR+D QG+RCGDMRPGLLRSEWE++
 5 Sbjct: 1 LHKIELGCDYDNKQSQAVARKLGFTLEANARDKRDVQGRKCDMRPGLLRSEWEKQ 58

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 230

- 10 A DNA sequence (GBSx0244) was identified in *S.agalactiae* <SEQ ID 733> which encodes the amino acid sequence <SEQ ID 734>. This protein is predicted to be ribosomal-protein-alanine N-acetyltransferase. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence
 15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4066 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 20

A related GBS nucleic acid sequence <SEQ ID 9599> which encodes amino acid sequence <SEQ ID 9600> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

- 25 >GP:BAB04418 GB:AP001509 ribosomal-protein-alanine
 N-acetyltransferase [Bacillus halodurans]
 Identities = 63/185 (34%), Positives = 95/185 (51%), Gaps = 11/185 (5%)
 Query: 53 KALPKLETDRLLRORTVGDVPAMFDYVCLSEVAYPAGLSPIASLEDEYDYFENRYQNL 112
 K P LET RLILR+ T D ++ Y+ +EV GL P +LED E +Y+++
 30 Sbjct: 6 KRPFILETKRLILRKITTDARSILSVLSDEKVMKYPGLEPFQTLLEDALG--ETIAYESI 63
 Query: 113 EKAKPSPGSGYITVKGSDRIIGSCAFN-----HRHEDDVFEICYLLHPDYWGHGYMTEAVA 167
 + +GIT+KG D +IGSC F+ H + FE+ L YWG G +EA+
 35 Sbjct: 64 LHEQTGIRWGITLKGQDEVIGSCGFHWVFKHRAEIGPELSKL----YWGQGIASEAIR 119
 Query: 168 ALIEVGFTLNLHKEIRCYDNKQSRVARKLGFTLEATIRDRKQNDNRCVNLIIYGLL 227
 A+I+ GF L L +I+ N S+R+ RK GF E +R + +Y LL
 Sbjct: 120 AVIQYGFHLLRLQIALIEPPNIPQRLVKEQGFISEGLLRSEYETTCGFDLLYMSLL 179
 40 Query: 228 RSEWE 232
 + +++
 Sbjct: 180 KRDFD 184

There is also homology to SEQ ID 732:

- 45 Identities = 39/54 (72%), Positives = 44/54 (81%)
 Query: 179 LHKIEIRCYDNKQSRVARKLGFTLEATIRDRKQNDNRCVNLIIYGLLRSEWE 232
 LHKIE+ CYDNKQS+ VA KLGFTLEA RDRKD Q RC ++ +GLLRSEWE
 50 Sbjct: 1 LHKIELGCDYDNKQSQAVARKLGFTLEANARDKRDVQGRKCDMRPGLLRSEWE 54

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 231

A DNA sequence (GBSx0245) was identified in *S.agalactiae* <SEQ ID 735> which encodes the amino acid sequence <SEQ ID 736>. Analysis of this protein sequence reveals the following:

```

Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2719 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 232

A DNA sequence (GBSx0246) was identified in *S.agalactiae* <SEQ ID 737> which encodes the amino acid sequence <SEQ ID 738>. Analysis of this protein sequence reveals the following:

```

Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3250 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9597> which encodes amino acid sequence <SEQ ID 9598> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 739> which encodes the amino acid sequence <SEQ ID 740>. Analysis of this protein sequence reveals the following:

```

Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3293 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 24/55 (43%), Positives = 38/55 (68%)

Query: 56 LLEEGITANKQDVLRKAGLVSLKAPAKVSSADVLALKEGIGPAATKQLVNKGVVFAK 110
      ++ G+ ++ + L G+ S +AF + +E D+LALKEGIGPA +K+LV+NG F K
Sbjct: 77 VVAGIRSDLVNITLYAEGIHSAQAFCWETKDLLALKGIGPATVKELVNGASPKK 131

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 233

A DNA sequence (GBSx0247) was identified in *S.agalactiae* <SEQ ID 741> which encodes the amino acid sequence <SEQ ID 742>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 25
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2901(Affirmative) < succ>
10     bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 743> which encodes the amino acid sequence <SEQ ID 744>. Analysis of this protein sequence reveals the following:

```

15     Possible site: 27
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2536(Affirmative) < succ>
20     bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

25     Identities = 57/84 (67%), Positives = 73/84 (86%)

      Query: 1  MSYBQEFLLKDFEEMWLSQOISINQMAMDSAKKVLSEDKDERAADAYIRYESKLDAYRFLOG 60
      MSYB+EFLKDFE+W+++QI +NQ+AM ++++V +ED DERA DA+IRYESKLDAY FL G
      Sbjct: 1  MSYBKEFLKDFEDWVKIQIQVQLAMATSQEVAQEDGERADAFIRYESKLDAYEFLIG 60

30     Query: 61  KFNINYENQKSFHDLPLDGLFGQRHY 84
      KF+NY N K+PHD+PD LFG RHY
      Sbjct: 61  KFMNYKNGKAFHIDPLFGARHY 84

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 234

A DNA sequence (GBSx0248) was identified in *S.agalactiae* <SEQ ID 745> which encodes the amino acid sequence <SEQ ID 746>. This protein is predicted to be methyltransferase. Analysis of this protein sequence reveals the following:

```

40     Possible site: 61
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2469(Affirmative) < succ>
45     bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 747> which encodes the amino acid sequence <SEQ ID 748>. Analysis of this protein sequence reveals the following:

```

50     Possible site: 35
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3352(Affirmative) < succ>

```

-302-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

5 Identities = 26/60 (43%), Positives = 37/60 (61%)

Query: 23 LKNERCPHFKLINVLEPKLEITLGDQKHILEKDSLISLSPQTHHLRAIENSKFLQIQLD 82
 + E P K+I VLE +L L DOK +L ++SLI++ Q+ HIL A + K LQ+ LD
 10 Sbjct: 42 ISQRTSPRDKVLTVLEGGQLFDLEDQKQVLTQESLIATPAQKVHLEARTDCKLLQVLLD 101

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 235

15 A DNA sequence (GBSx0249) was identified in *S.agalactiae* <SEQ ID 749> which encodes the amino acid sequence <SEQ ID 750>. This protein is predicted to be integrase (codV). Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3842 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

25 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 236

30 A DNA sequence (GBSx0250) was identified in *S.agalactiae* <SEQ ID 751> which encodes the amino acid sequence <SEQ ID 752>. Analysis of this protein sequence reveals the following:

Possible site: 22
 >>> May be a lipoprotein

35 ----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

40 No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 752 (GBS128) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 23 (lane 5; MW 15kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 32 (lane 4; 2 bands).

45 The GBS128-GST fusion product was purified (Figure 198, lane 2) and used to immunise mice. The resulting antiserum was used for FACS (Figure 288), which confirmed that the protein is immunoreactive on GBS bacteria.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 237

A DNA sequence (GBSx0251) was identified in *S.agalactiae* <SEQ ID 753> which encodes the amino acid sequence <SEQ ID 754>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2940 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 755> which encodes the amino acid sequence <SEQ ID 756>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2518 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 30/90 (33%), Positives = 49/90 (54%), Gaps = 10/90 (11%)

Query: 3   TVAVRVDQGLKIDATLFPQSLGLDMSTAVIMFLIQSVKTQSIPPEIK-----NKESSV 54
T+ +RVD +X A ++ + LG+ MST+ NPL Q + T IFF++ N +
Sbjct: 15  TLNLRVDDSVKSAADDILKRLGIPMSTAIMPLINQIILTGGIFPDVSLPEAPQRVNDYM 74

Query: 55  SDSEFQNLVETKLGIKIRKASDPESVNAFF 84
S E+P + + T + K +P+ V P+
Sbjct: 75  SQEKFYDKLITSPED--AKTNPQGVGKFY 102
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 238

A DNA sequence (GBSx0252) was identified in *S.agalactiae* <SEQ ID 757> which encodes the amino acid sequence <SEQ ID 758>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -2.81 Transmembrane 370 - 386 ( 368 - 388)

----- Final Results -----
bacterial membrane --- Certainty=0.2126 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9593> which encodes amino acid sequence <SEQ ID 9594> was also identified. A related GBS nucleic acid sequence <SEQ ID 10773> which encodes amino acid sequence <SEQ ID 10774> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 759> which encodes the amino acid sequence <SEQ ID 760>. Analysis of this protein sequence reveals the following:

```

Possible site: 37
>>> Seems to have a cleavable N-term signal seq.
      INTEGRAL    Likelihood = -4.57    Transmembrane    354 - 370 { 353 - 371)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.2826(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

15  LPXTG motif: 344-348

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 64/277 (23%), Positives = 99/277 (35%), Gaps = 31/277 (11%)

20  Query: 126 SIGNLFDLPKGGTTFVAFETFPVDIATPGDKPAKVVVTFDGSKDITVDVTVKVVDPRTDADKN 185
      ++ +LP ++ TT E PV ++ V + D+ + T P A
      Sbjct: 121 AVKDLPASTESTTQVPAFVQETQASASDSMTVGDSTSVTTDSPEETPSSESPVAFALSE 180

25  Query: 186 DPAGKDQQVNVGETPKAEDSIGNLFDLPKGGTTFVAFETFPVDIATPGDKPAKVVVTFDGSK 245
      PA Q E P S P T A ETP + A P P + S+
      Sbjct: 181 APA----QPAESEEPSVAASSEETPS--PSTPAAPETPEEPAAPSPSPSEEPSVAAPSE 234

30  Query: 246 DTVDTVTVKVVDPRTDADKNDPAGKDQQVNVGETPKAEDSIGNLFDLPKGGTTFVAFETFPVD 305
      +T P A + PA ++ T + P P + +TP
      Sbjct: 235 ETPSPET----PEEPAAPSPQPAESEEPSVAATTSPS-----PSTPAESET--QTTPAV 281

35  Query: 306 ATPGDKPAKVVVTFDGSKDITVDVTVKVVDPRTDADK-----NDPAGKDQQVNGK 355
      DKP+ P S + TV+ + +DK N + + +
      Sbjct: 282 TKDSKPPSSAAEK-PAASSLVSEQTVQQPTSKRSSDKKEQEQSYSPNRSLRQVRAHES 340

40  Query: 356 GNKLPATCGENATPPFNVALTINSVGLLSVSKKKED 392
      G LP+TGE A P P + +T+MS G L V+K+***
      Sbjct: 341 GKVLPSITGEKAQPLP-IATMTLMSLPSSLLVTKRQKE 376

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 239

A DNA sequence (GBSx0253) was identified in *S.agalactiae* <SEQ ID 761> which encodes the amino acid sequence <SEQ ID 762>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

```

Possible site: 49
>>> Seems to have no N-terminal signal sequence

50  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.5289(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 240

A DNA sequence (GBSx0254) was identified in *S. agalactiae* <SEQ ID 763> which encodes the amino acid sequence <SEQ ID 764>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -1.06 Transmembrane 39 - 55 ( 39 - 55)

----- Final Results -----
bacterial membrane --- Certainty=0.1426 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9591> which encodes amino acid sequence <SEQ ID 9592> was also identified.

The protein differs significantly from U58333 in several places:

```
Query: 157 TKPDGQVDIVNVSLTIYNSALRDKIDEVKK-----KAED-----PKNDEGSRDK 201
          T PDG D V+V++ ++ DK D K KAED P +G+
Sbjct: 683 TYPDGSKDTVDVTVKVVDPRITADKNDPAGKQQVNVGETPKARDSIGNLPDLPGKGTIVA 742

Query: 202 VLISLDDIKTDIDNNPK---TQSDIANKITEVTINLEKILVPRIPADKNDPAGKQQVNV 258
          +D T D K T D + +VT K++ PR DADKNDPAGKQQVNV
Sbjct: 743 FETFPVDTA-TPGDKPAKVVTYTPDGSKDTVDVT--VKVVDPRIT-DADKNDPAGKQQVNV 798

Query: 157 TKPDGQVDIVNVSLTIYNSALRDKIDEVKK-----KAED-----PKNDEGSRDK 201
          T PDG D V+V++ ++ DK D K KAED P +G+
Sbjct: 841 TYPDGSKDTVDVTVKVVDPRITADKNDPAGKQQVNVGETPKARDSIGNLPDLPGKGTIVA 900

Query: 202 VLISLDDIKTDIDNNPK---TQSDIANKITEVTINLEKILVPRIPADKNDPAGKQQVNV 258
          +D T D K T D + +VT K++ PR DADKNDPAGKQQVNV
Sbjct: 901 FETFPVDTA-TPGDKPAKVVTYTPDGSKDTVDVT--VKVVDPRIT-DADKNDPAGKQQVNV 956

Query: 157 TKPDGQVDIVNVSLTIYNSALRDKIDEVKK-----KAED-----PKNDEGSRDK 201
          T PDG D V+V++ ++ DK D K KAED P +G+
Sbjct: 288 TYPDGSKDTVDVTVKVVDPRITADKNDPAGKQQVNVGETPKARDSIGNLPDLPGKGTIVA 347

Query: 202 VLISLDDIKTDIDNNPK---TQSDIANKITEVTINLEKILVPRIPADKNDPAGKQQVNV 258
          +D T D K T D + +VT K++ PR DADKNDPAGKQQVNV
Sbjct: 348 FETFPVDTA-TPGDKPAKVVTYTPDGSKDTVDVT--VKVVDPRIT-DADKNDPAGKQQVNV 403

Query: 157 TKPDGQVDIVNVSLTIYNSALRDKIDEVKK-----KAED-----PKNDEGSRDK 201
          T PDG D V+V++ ++ DK D K KAED P +G+
Sbjct: 606 TYPDGSKDTVDVTVKVVDPRITADKNDPAGKQQVNVGETPKARDSIGNLPDLPGKGTIVA 663

Query: 202 VLISLDDIKTDIDNNPK---TQSDIANKITEVTINLEKILVPRIPADKNDPAGKQQVNV 258
          +D T D K T D + +VT K++ PR DADKNDPAGKQQVNV
Sbjct: 664 FETFPVDTA-TPGDKPAKVVTYTPDGSKDTVDVT--VKVVDPRIT-DADKNDPAGKQQVNV 719

Query: 157 TKPDGQVDIVNVSLTIYNSALRDKIDEVKK-----KAED-----PKNDEGSRDK 201
          T PDG D V+V++ ++ DK D K KAED P +G+
Sbjct: 446 TYPDGSKDTVDVTVKVVDPRITADKNDPAGKQQVNVGETPKARDSIGNLPDLPGKGTIVA 505

Query: 202 VLISLDDIKTDIDNNPK---TQSDIANKITEVTINLEKILVPRIPADKNDPAGKQQVNV 258
          +D T D K T D + +VT K++ PR DADKNDPAGKQQVNV
Sbjct: 506 FETFPVDTA-TPGDKPAKVVTYTPDGSKDTVDVT--VKVVDPRIT-DADKNDPAGKQQVNV 561

Query: 157 TKPDGQVDIVNVSLTIYNSALRDKIDEVKK-----KAED-----PKNDEGSRDK 201
          T PDG D V+V++ ++ DK D K KAED P +G+
```

-306-

Sbjct: 920 TYPGSKDTVDVTVKVVDPRTDADKMDPAGKQQQVNVGETPKAEDSIGNLPDLPGKITVA 979
 Query: 202 VLISLDDIRTDIDNNPK---TQSDIANKITVIMLEKILVPRIPDADKMDPAGKQQQVNV 258
 +D T D K T D + +VT K++ PR DADKMDPAGKQQQVNV
 5 Sbjct: 980 FETFPVDTA-TPGDKPAKVVVVTYPGSKDTVDVT--VKVVDPR-T-DADKMDPAGKQQQVNV 1035

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 241

A DNA sequence (GBSx0255) was identified in *S.galactiae* <SEQ ID 765> which encodes the amino acid sequence <SEQ ID 766>. This protein is predicted to be ara-C-like activator. Analysis of this protein sequence reveals the following:

15 Possible site: 30
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.37 Transmembrane 8 - 24 (8 - 25)
 ----- Final Results -----
 20 bacterial membrane --- Certainty=0.1150 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9589> which encodes amino acid sequence <SEQ ID 9590> was also identified.

25 There is homology to SEQ ID 460.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 242

A DNA sequence (GBSx0256) was identified in *S.galactiae* <SEQ ID 767> which encodes the amino acid sequence <SEQ ID 768>. Analysis of this protein sequence reveals the following:

30 Possible site: 53
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 35 bacterial cytoplasm --- Certainty=0.1200 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

40 A related GBS nucleic acid sequence <SEQ ID 9587> which encodes amino acid sequence <SEQ ID 9588> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 769> which encodes the amino acid sequence <SEQ ID 770>. Analysis of this protein sequence reveals the following:

45 Possible site: 50
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0679 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 135/176 (76%), Positives = 161/176 (90%)

```

5  Query: 1  MSYVYKDRQIQKTKVAIYNAPISILLQENDYSKITVQDVIGLANVGRSTFYSHYESKEVLL 60
    +S M KDRQI+KTK AIY+APT+LLQ+ +YSKITV+D+I LANVGRSTFY+HYESKE+LL
    Sbjct: 1  VSDMTKDRQIKKTKTAIYSAFTALLQKKYRKITVRDMITLANVGRSTFYAHYESKEMLL 60

10 Query: 61  KELCEDLFHHLFKQGRDVTTFREYLVHILKHFQKQKDSIATLLSDOPYLLFRFSELESD 120
    KELCE+LFHHLF+Q R+VTFE+YI+VHILKHFQK+DSIATLLS+DPYLLFRF+RLESD
    Sbjct: 61  KELCEELFHLFRQKHNVT7EDYLVHILKHFQKQKDSIATLLSDOPYLLFRFKNELSD 120

    Query: 121 VYPLRREYITKVDIPEDFLKQFILLSSFIETLKWLLHQKQMTVEDLLKYIITMVE 176
    VYP LR +YI K IFE FIAQF+LSSFIETLKWLLHQKQ+M+ +LLKYIL +++
15  Sbjct: 121 VYPNRCKYIDKTTIPEVFLKQFVLSSFIETLKWLLHQKQMSANELIKYILELIK 176
  
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 243

- 20 A DNA sequence (GBSx0257) was identified in *Sagalactiae* <SEQ ID 771> which encodes the amino acid sequence <SEQ ID 772>. Analysis of this protein sequence reveals the following:

Possible site: 29
>>> Seems to have no N-terminal signal sequence

- 25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3573 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 30 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 244

- 35 A DNA sequence (GBSx0258) was identified in *Sagalactiae* <SEQ ID 773> which encodes the amino acid sequence <SEQ ID 774>. Analysis of this protein sequence reveals the following:

Possible site: 35
>>> Seems to have a cleavable N-term signal seq.
 40 INTEGRAL Likelihood = -10.19 Transmembrane 112 - 128 (107 - 131)
 INTEGRAL Likelihood = -8.07 Transmembrane 77 - 93 (71 - 97)
 INTEGRAL Likelihood = -6.10 Transmembrane 144 - 160 (138 - 165)
 INTEGRAL Likelihood = -3.03 Transmembrane 165 - 181 (164 - 182)

- 45 ----- Final Results -----
 bacterial membrane --- Certainty=0.5076 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 775> which encodes the amino acid sequence <SEQ ID 776>. Analysis of this protein sequence reveals the following:

Possible site: 35
>>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -9.13 Transmembrane 112 - 128 (107 - 130)

-308-

INTEGRAL	Likelihood = -5.89	Transmembrane	144 - 160 (138 - 163)
INTEGRAL	Likelihood = -5.47	Transmembrane	7 - 23 (6 - 29)
INTEGRAL	Likelihood = -3.50	Transmembrane	77 - 93 (74 - 94)
INTEGRAL	Likelihood = -2.07	Transmembrane	166 - 182 (165 - 183)

----- Final Results -----

bacterial membrane --- Certainty=0.4652(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 212/287 (73%), Positives = 245/287 (84%)

```

Query: 1  MTSNKKVAIAFILNISFVSIFSGSLFFSGAILADAVHDPGDAIAIGISATLEKKSCKD 60
      M ++KKV I FILN+SFS++BFIFG+LFFSGAILADAVHDPGDAIAIGISA LE+K+ K
Sbjct: 1  MPASKKVTIIFILNLSFSLBFIFGTLLFFSGAILADAVHDPGDAIAIGISAILERKAVKK 60

Query: 61  EDTFISLGYKRFSLLGADITSLILISGSLVMIENIPKLMHPTVFNVHGMFLAVIAIII 120
      E FSLGYKRFSLLGAL T+LILISGS+LVMI IPKLMHPT VNY GMF+LA+ AIII
Sbjct: 61  ESPNFSLGYKRFSLLGALTNLILISGSLVMIETIPKLMHPTIVNYDGMFVLAIFAIII 120

Query: 121  NGLASFILHSGQSKHEILLSLHFLIEDILGWLAIIVISLILNWKPLFYILDPLLSVAISTFI 180
      NG ASPI+HS Q+K-EILLSLHFLIEDILGWLAII++SLIL WKP YILDPLLS+AI++FI
Sbjct: 121  NGFASFTIHSNQTREEILLSLHFLIEDILGWLAIITLSLILNKKPWIYILDPLLSIAIASFI 180

Query: 181  LSKALPKLLSTLKLFLDGVFSDIDYAAHLDELKGLSQVRSINGLNHWSMDGIDNRAIHC 240
      LSKALPKL++T +FLDGVFSDIDY LH EL L + S+NQLN+WSMDGID+RA IHC
Sbjct: 181  LSKALPKLVATAN:FLDGVFSDIDYCTLHHELSQLPHIVSVNQLNWSMDGIDIRATHC 240

Query: 241  CLNQLISEKDCRAITICQHYKINDVTVIEDISLREHQNHCKPLQN 287
      CL + +EK CK++IR IQY Y IN VIVEID SL EHQ+RC L +
Sbjct: 241  CLRSTTEHCKKSIKILIQRNINSVTVIEDTSLNHHQHCCSSLS 287

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 245

A DNA sequence (GBSx0259) was identified in *S.galactiae* <SEQ ID 777> which encodes the amino acid sequence <SEQ ID 778>. Analysis of this protein sequence reveals the following:

Possible site: 48
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -1.22 Transmembrane 221 - 237 (221 - 237)

----- Final Results -----

bacterial membrane --- Certainty=0.1489(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

There is also homology to SEQ ID 780.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 246

A DNA sequence (GBSx0260) was identified in *S.galactiae* <SEQ ID 781> which encodes the amino acid sequence <SEQ ID 782>. Analysis of this protein sequence reveals the following:

Possible site: 13
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -2.50 Transmembrane 2 - 18 (1 - 18)

-309-

----- Final Results -----

bacterial membrane --- Certainty=0.1999 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 247

A DNA sequence (GBSx0261) was identified in *S.agalactiae* <SEQ ID 783> which encodes the amino acid sequence <SEQ ID 784>. This protein is predicted to be dehydrogenase (Zn-dependent). Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -3.77 Transmembrane 171 - 187 (170 - 187)

----- Final Results -----

bacterial membrane --- Certainty=0.2508 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAQ20655 GB:AE005134 alcohol dehydrogenase; Adh2 [Halobacterium

sp. NRC-1]

Identities = 169/348 (48%), Positives = 232/348 (66%), Gaps = 9/348 (2%)

Query: 1 MKVATFIEPGKRVITITPKPVIEQETDAVIKIVRACVCGSLDWYRGISKRESGSPAGHE 60
 M+ A + PG++ + + PKP IE DAVI++ VCGSLDW+YRG S RE+GS GHE
 Sbjct: 1 MRAAVYQGPGEIAVEEVPKPDIESPEDAVIRVTRTAVCGSLDWYFGSDSDREAGSRVGHE 60

Query: 61 AIGIVEVGTKVTDVSKGDFVIVPFTTHGCGQCPCKAGFDGNCNTNHA--AAKVGYGQ 117
 +GIVEVG VT V+ GD VI PF CG+C C+ G +C ++ N G QG+
 Sbjct: 61 FMGIVEVEGDDVTSVANODRVIAFPFISCGBCFPCROGLYTSCEVEDSWGSRANGGGQGR 120

Query: 118 YLRYNANWALVKIPGQPSDYINETIANSL/LSIDVMAATGYHAATAEVRKEDTVVVGDG 177
 Y++ A+ LV++P + +D D + L SIL L+DVM TG+HAA +A V RGDV VV+GDG
 Sbjct: 121 YVICKCFADGTLVRVPDRYAD-DEVDLSLEPL/LTDVMTGHPAAVAGVGBGDVAVVVGDG 179

Query: 178 AVGLCGVIAAKHMLGANRIIAMSRRHKRQELATFGATDIVEERGEDAIVKRVLDLTNCGA 237
 AVGLCGV+AA+ LGA RIIM H+DR ELA FGATD + RGD+H++R DLT+ GA
 Sbjct: 180 AVGLCGVLAQRKLGABRIIANGHEDRIELAAEPGATTTISARGHDAERARDLTH-GGA 238

Query: 238 DAVLECVGTQBSVDITATQIARPGAVIGRVGIP--QNPDMNTNNLWKNIGLGGIASVT 294
 +V+ECVG ++D+A IARPG +G VG+P ++ ++ +F NI +RGG+A V
 Sbjct: 239 NHVMECVGAASAMDSAIATARPGETVGYGVPGVREDGGELVPTMTFSDNITTRGVA PVR 298

Query: 295 TFDKSVLLDAVLTHKINPLVPTKSVFLDDIQKAYEAMDKRDAIKSLV 342
 + + ++ D VL ++P +PTK+ LD + + Y AMD R+AIK LV
 Sbjct: 299 AYAEELNAD-VLQGTLDPSPIFTKTVLDDGVPRGYAAMDREAIKVLV 345

There is also homology to SEQ ID 786.

A related sequence was also identified in GAS <SEQ ID 9145> which encodes the amino acid sequence <SEQ ID 9146>. Analysis of this protein sequence reveals the following:

Possible site: 23

- >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -5.41 Transmembrane 170 - 186
- 5 ----- Final Results -----
 bacterial membrane --- Certainty=0.3166 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
- 10 An alignment of the GAS and GBS proteins is shown below:
 Identities = 121/353 (34%), Positives = 182/353 (51%), Gaps = 16/353 (4%)
- Query: 1 MKVATFIEPKGMVITDTPKPVIEQETDAVIKIVRACVCGSLMWYRG-ISKRRSGSPAGH 59
 15 MK AT++ G + + D PKFVI + TDA++++V+ +CG+DL G + + G+ GH
 Sbjct: 15 MKAATYLSYGNLQLIDKPKFVILKPTDAIVQLVKTTICQTLDLHLGGDVPAKCEGTILGH 74
- Query: 60 BAIGIVEVGTKVTVDSKGFVIVPFTHGCGQCPSCKAGFDGNCNTNHQAANK---VGYQG 116
 E IGIV+EVG VT+ GD VI+ C C CK G +C + G Q
 20 Sbjct: 75 BGGIVKVEVGDVTFNFKIGDKVIIISCVTSCHTCYCKRGLSSHCCQDGGWILGHLLNTQA 134
- Query: 117 QYLRYTNANMALVKIPQSPDYDNETLNSLLTLDVMTAGYH-AAATABVRKEDTVVVMG 175
 +Y+ ++ +L P D +L+ LSD+ + T Y + VR GD V ++G
 25 Sbjct: 135 KYVHLPHADGSLYHAPDTDD-----KALVMSDILPFTSYEIGVLPFHVKGNDVCIVG 188
- Query: 176 DGAIVGLGVIAAMLANRIIAMSRRHKRQELATFTGATDIVEERGDEAVKRVL-DLTNQ 234
 G VGL ++ + II + ++R E A TFGAT + E VK ++ D+TN
 30 Sbjct: 189 AGPVGLAALLTVQFFSPANIMVDSQNRLEAAKTFGATHICSGSSSEVKALIDDTING 248
- Query: 235 AGADAVLECGVTGQSDVDTATQIARPGAVIGRWGPQNP-DMNTNLFWNIGLRGGIASV 293
 G D +ECVG + D +I G I VG+ P D N + L+ KNI L G+ +
 35 Sbjct: 249 RGVDLSMBGVGPATFDICQKIIISVGGHIANVGVHGKPVDFNLDELWIKNTILNTLNTLVNA 308
- Query: 294 TTFDSVLLDAVLTHKINPGLVFTKSFVLDDIKAYEAMDKRDAIKSL-VIVD 345
 T + +LL+ + T KI+ + T F L +++KAYE A +L VI+ +
 35 Sbjct: 309 NTTE--MLNLVKTGKIDATRLTHHFKLSEVEKAYETFKHAGANNAKLVVID 359

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 248

- 40 A DNA sequence (GBSx0262) was identified in *S.agalactiae* <SEQ ID 787> which encodes the amino acid sequence <SEQ ID 788>. Analysis of this protein sequence reveals the following:
 Possible site: 31
 >>> Seems to have no N-terminal signal sequence
- 45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2169 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
- 50 The protein has homology with the following sequences in the GENPEPT database:
 >GP:AAD36075 GB:AB001762 hypothetical protein [Thermotoga maritima]
 Identities = 55/128 (42%), Positives = 72/128 (55%), Gaps = 8/128 (6%)
- 55 Query: 8 IFPKGEKNPYGEFFIGQSYLAALAKSPDC--NVSGVNTVFAGCRNRMWHVILGDYQILLV 65
 IF +G K +FF G ++ L +G N V +V FE G R +WH H G QIL+V
 Sbjct: 5 IFRGSKGS-SDFTGNVWVKMLVTDNRGVFTQVDVVFEPGARTWHSHPGG-QILLV 62
- Query: 66 TEGSGWYQEEGKAVSLKPGDVIVIDKGVRRHWRGAQKQSEFAHIAITA---GKSEFYFA 121
 T G G+YQK GK A LK GDV+ V HWHA D E HI I+ G +E+ +
 60 Sbjct: 63 TRGKSFYQERKPARILKKGDVVEIPANVVRHWRGAAPDEELVHIGISTQVHLGPAEWLGS 122

-311-

Query: 122 VSDDEYSR 129
 V++REY +
 Sbjct: 123 VTESEYRK 130

- 5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 249

- A DNA sequence (GBSx0263) was identified in *S.agalactiae* <SEQ ID 789> which encodes the amino acid sequence <SEQ ID 790>. This protein is predicted to be gamma-carboxymuconolactone decarboxylase. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4089 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 20 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA20070 GB:AL031155 3-oxoadipate enol-lactone
 hydrolase/4-carboxymuconolactone decarboxylase
 [Streptomyces coelicolor A3(2)]
 Identities = 33/93 (35%), Positives = 59/93 (62%), Gaps = 1/93 (1%)

25 Query: 11 QLEEFAPDEFARYNDLLFGKVMKEDHLTDKTRSIITISALISGGNLEQLEHHQFAKQN 70
 Q +EF+ +P + +GB+W + L ++RS +T++hL++GG+L++L HL+ A +N
 Sbjct: 349 QADEFSGDFQEFILTRYANGKIDRPG-IDRRSRSCVTLLALVAGGHLDELAPHLRAALRN 407

30 Query: 71 GVTKEETADITHTLAFYVGNPIAMSAFNKAKEI 103
 G+T EI +++ A Y G P A AF A+++
 Sbjct: 408 GLTPGEIKEVLLQAAYVGVPAANGAFKVAQQV 440

No corresponding DNA sequence was identified in *S.pyogenes*.

- 35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 250

A DNA sequence (GBSx0265) was identified in *S.agalactiae* <SEQ ID 791> which encodes the amino acid sequence <SEQ ID 792>. Analysis of this protein sequence reveals the following:

- 40 Possible site: 44
 >>> Seems to have no N-terminal signal sequence
- 45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5529 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 251

A DNA sequence (GBSx0266) was identified in *S.agalactiae* <SEQ ID 793> which encodes the amino acid sequence <SEQ ID 794>. This protein is predicted to be probable transcriptional regulator. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9585> which encodes amino acid sequence <SEQ ID 9586> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAG08263 GB:AE004901 probable transcriptional regulator
[Pseudomonas aeruginosa]
Identities = 36/148 (24%), Positives = 68/148 (45%), Gaps = 22/148 (14%)

Query: 5 QIVEKPAFLLAG-----VTLENVKSNGBQICQAIGICKTQPDPRFD 45
+IVE+PA + G + E+ + + + GIC QP+ F
Sbjct: 123 RIVERPAFSPVVGMEYFGSAPGDTIGQWIERFIPREHRLAGKHDPRVSYGICACQPNGEFH 182

Query: 46 YGATYQVETSVQAPKGLIIRIPSATYAVISVKGPMPSLQETWRKLIQGFQENNLKPA 105
Y A ++V+ P+G+ ++P+ YAV + KG P + E+++ I E L+P
Sbjct: 183 YVAGFEVQEGWPVPEGMVRFQVPAQKYAVFTHGTAP-QIARSPQAIYSHLAEERGLEPK 241

Query: 106 NSPFLNEYSSQH--PQDTDYQMEINLAI 131
+ E Y + P D + Q++++ I
Sbjct: 242 AGVDFFEYDQRFGRGLDPNSQVDLYIPI 269
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 252

A DNA sequence (GBSx0267) was identified in *S.agalactiae* <SEQ ID 795> which encodes the amino acid sequence <SEQ ID 796>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0887 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB84919 GB:AE000825 conserved protein [Methanothermobacter
thermautotrophicus]
Identities = 42/130 (32%), Positives = 71/130 (54%), Gaps = 3/130 (2%)
```

-313-

Query: 1 MITQEMKEIINSCLAMVATVDAGKQFNIGPKRSMRLNDKTFIYMENTDQQTINIEDNG 60
 K+T EM + I +L VAT D +G FN+ P R D+T + +N +T N+ +N
 Sbjct: 1 MMTPEMDAIEKEKLVFVATDEBGTNNVVPIDFARPLDSRTILIDNMYKKKTIRNLRNP 60

5 Query: 61 KIEIAFVDRERLLGYRFVGTABICTETGYTAAKKNWGRMG--VPKVGIIHVERIFNL 118
 +I + R Y+P GT EI G Y++ +WA+ M PK+ ++ VE I+++
 Sbjct: 61 RIAL-IPQARECTPYCFKGTVEIFPGKGYFDMVVEWQVMMTELEPEISAILMTVERITYSV 119

10 Query: 119 QSGANAGKEI 128
 + G AG+++
 Sbjct: 120 KPGPEAGEKV 129

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 797> which encodes the amino acid sequence <SEQ ID 798>. Analysis of this protein sequence reveals the following:

15 Possible site: 24
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 20 bacterial cytoplasm --- Certainty=0.0789 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 123/128 (96%), Positives = 127/128 (99%)

25 Query: 1 MITQEMKEIINSCLAMVATVDAGKQFNIGPKRSMRLNDKTFIYMENTDQQTINIEDNG 60
 MITQEMK++IN+CLAMVATVDAGKQFNIGPKRSMRLNDKTFIYMENTDQQTINIEDNG 60
 Sbjct: 1 MITQEMKDLINCLAMVATVDAGKQFNIGPKRSMRLNDKTFIYMENTDQQTINIEDNG 60

30 Query: 61 KIEIAFVDRERLLGYRFVGTABICTETGYTAAKKNWGRMGVPKVGIIHVERIFNLQS 120
 KIEIAFVDRERLLGYRFVGTABICTETGYTAAKKNWGRMGVPKVGIIHVERIFNLQS 120
 Sbjct: 61 KIEIAFVDRERLLGYRFVGTABICTETGYTAAKKNWGRMGVPKVGIIHVERIFNLQS 120

35 Query: 121 GANAGKEI 128
 GANAGKEI
 Sbjct: 121 GANAGKEI 128

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 253

A DNA sequence (GBSx0268) was identified in *S.agalactiae* <SEQ ID 799> which encodes the amino acid sequence <SEQ ID 800>. Analysis of this protein sequence reveals the following:

Possible site: 22
 >>> Seems to have a cleavable N-term signal seq.
 45 INTEGRAL Likelihood = -5.47 Transmembrane 1028 -1044 (1027 -1048)

----- Final Results -----
 bacterial membrane --- Certainty=0.3187 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 50 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

55 !GB:AF054892 surface antigen BspA [Bacteroides forsy...
 !GB:AF054892 surface antigen BspA [Bacteroides forsy...
 !GB:AF054892 surface antigen BspA [Bacteroides forsy...
 !GB:AF054892 surface antigen BspA [Bacteroides forsy...
 !GB:AF054892 surface antigen BspA [Bacteroides forsy...
 >GF:AAC82625 GB:AF054892 surface antigen BspA [Bacteroides

for sythus]

Identities = 143/566 (25%), Positives = 243/566 (42%), Gaps = 52/566 (9%)

Query: 95 VPKAKPEVTVQRASNSNDASKVVEPKQDTASKKETLETSTWEAKLFVTRGDTGLV---F 150
+P + + + A + + + P TA + L T + T + G F

5 Sbjct: 120 PINSVTTIGEWAFKGCGLKSITLPSNLTAIGCSALSGCTGLTSITIPNSVTTIGEWAF 179

Query: 151 SKSGINKLSQTSHLVLPSEHAA--DGTQLTVQVASFPTDPKKTATAEVTSRLGEMGKPSRL 208
SG+ ++ + L +A LT+ PD TIE + G+G S

10 Sbjct: 180 GCSGLTSITFPNSLTAIGESAFYGCGLTSIT---LPDALTTIGESAFK-GCSGLKSIT 234

Query: 209 DIDQKEIIDBGEIENAYQLTKLTIPNGVKSIGQDAFVDMHKNIAEVNLPLESLETISDYAPA 268
+ I E ++ LT +T+P+ +IG+ AF + + P SL TI + AF

15 Sbjct: 235 FPNSLTTIGESAFYGCGLTSITLPDALTTIGESAFYGCGLKSITFPNSLTTIGESAF 294

Query: 269 HM-SLKQVKLPEDNLKVGIGELAFFNQIGGKLYLPHRLIKLAERAFKSNRIQTVEFLGSKL 327
+ SL + +P+++ IG AF+ + LP L + ERAF + T + + +

20 Sbjct: 295 NCGSLTSITIPNSVTTIGESAFYGCGLKSITLPGDLTTIEERAFYNCVGLTSITIPNSV 354

Query: 328 KVGESAFOD-NLNRNVMLPDGLEKIBSEAFITNGDEHYNNQVLRTRTQGNPHQLATE 386
IGE++F + L+++ LPDGL IE AF N L + T N E

25 Sbjct: 355 ATIGESAFYGCGLKSITLPGDLTTIEWGAFY-----NCGALTSITIPNSVTTIGES 405

Query: 387 NTVPNDPKSLWRATPDMQDYTKWLEDEFTYQKNVTFGS---NKGQLQVRNKNKLEIPKQ 443
+ + +L T D ++D +++ ++G G + V K ++K+

30 Sbjct: 406 SAFYGCGLKDWTVVMDTPIIDQD-VFRELTGSGIRLHVPAKGTQVYEAK-DWMEK 461

Query: 444 NGITITIGEDNAFNVDPQSKTLKLYDLSEIKLESTIRKIGAFAPQSNLKSFEASDELE 503
N + + G + ND +KTL + P T + + FA ++L

35 Sbjct: 462 NIVEDDDPGGLGN-NYDAATKILITIN---PTPDTPKMFNFATPDQLW----- 507

Query: 504 EIKGEAFMNNRIGTLDLKDLKIGDAAPH-INHIYAIVLPESVQIGRSAPFQNGALHL 562
GAF I + + D + +GD AF + + + I LP+SV IG+SAF L

40 Sbjct: 508 ---GAFQKQ-IQKITIGDGVTSVGDFAFGSCDALKSITLPSKVTITGQSAFSCDWRKS 562

Query: 563 MFIGNKVKITIGEMAFLSNKLSEVLSSEKQKLTIEVQAFS-DNALSEGVFPNLQITIRE 621
+ + + V TIGE AF + LE +++ K + I + F +L+ + LP L I ++

45 Sbjct: 563 LTLPGVNTIGEKAFY-DCELTSTITPKSVTAIGQSTPHYCVSLTSLTLPDALTAIGK 621

Query: 622 AF-KRNHLKIEVKGSTLSQITFNADF 646
AF N L V +++ I NADF

50 Sbjct: 622 AFYSCNALTSVTFPKSITTIGENADF 647

Identities = 109/407 (26%), Positives = 175/407 (42%), Gaps = 48/407 (11%)

Query: 222 FNAYQLTKLTIPNGKSIQGDAPFVDMKNIAEVNLPLESLETISDYAFANMS-LKQVKLPDN 280
P+ LT +T+PN +IG AF + + +P G+ TI ++AF S LK + LP++

55 Sbjct: 87 FSDCALTSVTLPSNLTAIGDHAFKGCGLTSITIPNSVTTIGEWAFKGCGLKSITLPMG 146

Query: 281 LKVGIGELAFFNQIGGKLYLPHRLIKLAERAFKSNRIQTVEFLGSKLVKIGESAFOD-NN 339
L IG+ A + +P + + E AF T + L IGE++F

60 Sbjct: 147 LTAIGCSALSGCTGLTSITIPNSVTTIGEWAFKGCGLTSITFPNSLTAIGESAFYGCGL 206

Query: 340 LRNVMLPDGLEKIBSEAFITNGDEHYNNQVLRTRTQGNPHQLATEINTVPNDKSLWRA 399
L ++ LPD L I AP G G L++ T N E+ + +

65 Sbjct: 207 LTSITLPDALTTIGESAFKGCGLS-----LKSITFPNSLTTIGESAFYGCGLTSIT 257

Query: 400 TPDMDYTKWLEDEFTYQKNVTFGSNKGQLQVRNKNKLEIPKQNGITITIGEDNAFNV 459
PD ++T K++ P ++T IG++AF N

70 Sbjct: 258 LPD-----ALTTIGESAFYGCGLKSITFPN-----SLTTIGESAFYNC 296

Query: 460 DPQSKTLKLYDLSEIKLESTIRKIGAFAPQS-NNLKSFEASDELEIKGEAFMNNRIGT- 517
L I +P+++ IG AF + LKS + L I+E AF N + T

75 Sbjct: 297 G-----SLTSITIPNSVTTIGESAFYGCGLKSITLPGDLTTIEERAFYNCVGLTS 347

Query: 518 LDKKDLKIGDAAPH-INHIYAIVLPESVQIGRSAPFQNGALHLFMIGNKVKITIGEMA 576
+ + + + IG++AF+ + + + I LP+ + I AF GAL + I N V TIGE A

80 Sbjct: 348 ITIPNSVATIGESAFYGCGLKSITLPGDLTTIEWGAFYNCVGLTSITIPNSVTTIGESA 407

Query: 577 FLS-NKLESVNLSEQQLKTIEVQAFSDNALSHVL--PPNLQTIIR 620
 F L+ V ++ I+ F + LS + L P + +T+ E
 Sbjct: 408 FYGCGALKDVTVMADTPI-DIQDIVFRELTLGIRLHVPAKKTIVR 453
 Identities = 111/465 (23%), Positives = 185/465 (38%), Gaps = 56/465 (12%)

5 Query: 141 VTRGDTLVGFSKSGINKLSQTSNHLVLPASHADGTQLQVASF- 190
 +T D L L +S S + P+ LT + AF PD T
 Sbjct: 210 TLPDALITIGESAFKGCGLKSITFPN-----SLTTIGESAFYDGCALTSITLPDALIT 263

10 Query: 191 ALASYTSRLGEMKPKRLDIDQKRIIDEGRIFNAYQLTKLTIIPNCKSIGQDAVFNIN 250
 I ++ G +G S + I E +N LT +TIPN +IG+ AF +
 Sbjct: 264 TIGR-SAFYGCGLKSITFPNSLTITIGESAFYNGSLTSITIPNSVTIGESAFYGCGL 322

15 Query: 251 AEVNLPESELETISDYAFHMS-LKQVKLPUNLKVIGELAFDNQIGGKLYLPHLKLK 309
 + LP+ L TI + AF + L + P+ ++ IGR AF+ + LP L +
 Sbjct: 323 KSLITPDGLTTIERAFYNGVLTSTITIPNSVATIGESAFYGCGLKSITLPDGLTIT 382

20 Query: 310 RAFKSNRIQTVEFLGSKLVIGESAFD-NHLRNVLHP-DGLEKIESAF-----TGD 362
 AF + T + + + IGE++F L++V + D I+ + F +G
 Sbjct: 383 GAFYNGCALTSITIPNSVSTIGESAFYGCGLKDVTVAMDTPIDIQRDVRELTLGIRL 442

25 Query: 363 DEHYNNQVVLRTTRQNPQLATKN-----TYVNDKSLMRATPMDYTKWLEEDPT 415
 + V + T + + + K+L P D K + +F
 Sbjct: 443 HVPAGKKTIVYKADVNKEFIVEDDPGSLQWYDAATKTLTITNPDPKPM-PKAT 501

30 Query: 416 QKISVTGFSNKGLOKVRNKNLEIPKQHNQITITIGMIAFRNVDPQSKTLAKVDLEIK 475
 + + G K +K+ G +T +GD AF D L+ I
 Sbjct: 502 PNDQLGAFQKZIKIT-----ICDGVTSVGDPAFSGCD-----ALKST 541

35 Query: 476 LPSTRIKIGAFAPQS-NLKSFEASDLEEKKGAPMN-NRIGTLDLKDILKIGDAAPH 533
 LP ++ IG AF +L+S + + I E AF + + + + IG FH
 Sbjct: 542 LPKSVTTIGQSAFSGCDLRSITLPDGVNTIGEKAFYDCELTITITIPKSVTAIGQ 601

40 Query: 534 -INHIAIVLPESVQIGRSAPFGNGLHAFIGNKVTIGENAF 577
 + ++ LP+++ IG+ AF AL + + TIGE AF
 Sbjct: 602 YCVSLTSLTLPDALTAIGKAFYSNALTSVTFPKSITTIGENAF 646
 Identities = 98/351 (27%), Positives = 152/351 (42%), Gaps = 53/351 (15%)

45 Query: 315 NRIGTVEFLGSKLVIGESAFQDNHLRNVLHPDGLEKIESAFPTNGDEHYNNQVVLRT 374
 ++IQTV +G + +G +F D L +V LP+ L I AF G G L +
 Sbjct: 68 SKIQTVI-IGDGVTSVGNKAFSDCALTSVTLNSLTAIGDHAFKGCSC-----LTS 117

50 Query: 375 RTGQNPOLATENTYVNDKSLMRATPMDYTKWLEEDPTQKNSVTGFSNKGLOKVRN 434
 T P+ + T + S ++ NS+T L
 Sbjct: 118 IT--IPNSVTTIGENAFKGCGLKSIT-----LPNSLTAIGQSALSGCTGL 161

55 Query: 435 KMLEIPKQNGITITIGENAF-----RNVDQSKTLKYD-----LEELKPSIT 480
 ++ IP ++T IG+ AF ++ F + L I LP +
 Sbjct: 162 TSITIPN-----SVTTIGENAFPGCGLTSITFPNSLTAIGESAFYGCGLTSITLPDAL 216

60 Query: 481 RKIGAFAPQS-NNLKSFEASDLEEKKGAPMN-NRIGTLDLKDILKIGDAAPH-INHI 537
 IG AF+ + L+S L I E AF + + + L D L IG +AF+ +
 Sbjct: 217 TTIGESAFKGCGLKSITFPNSLTTIGESAFYDGCALTSITLPDALTTIGRSAPYGCGL 276

65 Query: 538 YAVILPESVQIGRSAPFRONGALHFMIGNKVTIGENAF-LS-NKLESVNLSEQQLKTI 595
 +I P S+ IG SAF G+L + I N V TIG AF + L+S+ L + L TI
 Sbjct: 277 KSITFPNSLTITIGESAFYNGSLTSITIPNSVTIGESAFYGCGLKSITLPD--GI/TTI 334

Query: 597 EVQAFSD-NALSEVVLPPNLQTIIRKAPKR-NHLKRVKSSSTLQITFNAF 645
 E +AF + L+ + P ++ TI E AF + LK + L I + AF
 Sbjct: 335 ERERAFYNGVLTSTITIPNSVATIGESAFYGCGLKSITLPDGLTITBWAF 385
 Identities = 78/282 (27%), Positives = 123/282 (42%), Gaps = 46/282 (16%)

Query: 111 NDASKVEVPPKQDTASKKETTLETSWEARDPVTRGTLVGFSGKINKLSQTSNHLVLP 168
 N+AS E+P SK +T VT GD + + + + TS + LP+
 Sbjct: 56 NNAS--EIPHSLQSKIQT-----VTIGDGVTSVGNKAFSDCALTS-VTLPSL 101

Query: 169 -----HAADG-----TQLQVASFAT-----PKKTAIAETTERIGENG 203

-316-

HA G +T +AF P+ TAI + + G G
 Sbjet: 102 TAIGDEAFKCGSLTSITIPNSVTTIGEWAFKCGSLKSLTIPNSLTAIGQ-SALSGCTG 160

5 Query: 204 KPSRLDIDQKEILDGEIFNAYQLTKITINGYKSGIGQDAFVNKILAEVNLPESELETIS 263
 S + I E F LT +T EN +IG+ AF + + LP+L TI
 Sbjet: 161 LTSITIPNSVTTIGEWAFKCGSLTSITIPNSLTAIGESAFYGCAGLTSITLPDALTTIG 220

Query: 264 DYAFAIMS-LKQVKLFENLKVIGELAPFDMQIGCKLYLPRHLIKLAERAFKS-NRIQTVE 321
 + AF S LK + P+L IGE AF+D + LP L + AF + +++
 10 Sbjet: 221 ESAFKCGSLKSLTIPNSLTTIGESAFYDCCALTSITLPDALTTIGESAFYGCGLKSLT 280

Query: 322 FLGSLKLVIGESFQD-NLNRNVMLPDGLKEIIESEAFQNG 362
 F S L IGE+P + +L ++ +P+ + I AF G G
 Sbjet: 281 FPNS-LTTIGESAFYNGSLTSITIPNSVTTIGRSAPFYGCSG 321

15 Identities = 43/144 (29%), Positives = 70/144 (47%), Gaps = 4/144 (2%)

Query: 220 EIFNAYQ--LTKLTPNGYKSGQDAFVNKINIAEVNLPESELETISDYAFAIM-SLKQVK 276
 +++ A+Q + K+TI +G S+G AF + + LP+S+ TI AF+ L+ +
 20 Sbjet: 505 QLWGAPOKEIQITIGDGVTSVGDFAFSGCDALKSLTLPKSVTTIQSGAFSCWDLRLSIT 564

Query: 277 LPDNLKVIGELAFFNIGGKLYLPRHLIKLAERAFKSNRIQTVEFLGSLKLVIGESFQ 336
 LPD + IGE AF+D + +P+ + + F T L L IG+ +F
 Sbjet: 565 LPDGVTTIGESAFYDCLSLTSITIPKSVTAIQOETHYCVSLTSLPDLTAIGKAFY 624

25 Query: 337 D-NLNRNVMLPDGLKEIIESEAFQ 359
 N L +V P + I AF G
 Sbjet: 625 SCNALTSVTFPKSITTTIGENAFDG 648
 Identities = 43/134 (32%), Positives = 66/134 (49%), Gaps = 12/134 (8%)

30 Query: 511 MNNRIGTLIDKDKLIKIGDAAFHINHIAIVLPESVOEIGESAFRONGALHAMFIGNKVK 570
 + ++I T + D + +G+ AF + ++ LP S+ IG AF+ L + I N V
 Sbjet: 66 LQSIQTVTIGDGVTSVGNNAFSDCALTSVLPNSLTAIGDHAFKCGSLTSITIPNSVT 125

Query: 571 TIGEMAFS-NKLESVNLSEBQKQKLTIEVQAFSD-NALSEVVLPPNLQITRERAFKRNHL 628
 TIGE AF + L+S+ L L I A S L+ + +P ++ TI E AF
 35 Sbjet: 126 TIGEWAFKCGSLKSLTIPNSLTAIGQSAISGCTGLTSITIPNSVTTIGEWAF---- 178

Query: 629 KEVKGSSLSQITF 642
 G S L+ ITF
 40 Sbjet: 179 ---FGCSGLTSITF 189

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 801> which encodes the amino acid sequence <SEQ ID 802>. Analysis of this protein sequence reveals the following:

45 Possible site: 21
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -2.44 Transmembrane 984 -1000 (984 -1001)

50 ----- Final Results -----
 bacterial membrane --- Certainty=0.1977 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

55 LEXTG motif: 975-979

An alignment of the GAS and GBS proteins is shown below:

Identities = 751/1050 (71%), Positives = 861/1050 (81%), Gaps = 45/1050 (4%)

60 Query: 3 KKHLLTALALTTVSVVTVSQEVYGLERRESVKQEQTQSA-SRDWFEEDNERKTNVSK 61
 KKHLLT+AL LTTVSVVT++QEV+ L +E +KQ Q S+ S D+ E + K + ++
 Sbjet: 2 KKHLLTALITLTTVSVVTHQEVFLVKEPIKQTAQSSISGADYAESGSKKIKINET 61

Query: 62 NSTVDETVDLFDGNSNNSSKRTESVVDPKQVFKAPEVTOBASNSNDKAVKVPKQ 121
 + VD+TV+DLFSD + K +Q KA E T E+ S++E K+
 65 Sbjet: 62 GSPVDOTVIDLFDKRTTPEKIKDNLAKGPRBQELKAVTENT-BSEKQITGGSOI+QSK 120

5	Query: 122	DTASKKETLETSTWEADPFVIRGDTLVGFSSKSGINKLSQTSHTVLVPSHAADGTQLTQVAS	181
	Sbjct: 121	+ K TS WE DF+T+G+TLVG SKSG+ KLSQT HLVLPS AADGTQL QVAS	180
	Query: 182	FAFTPDKKTAIAEYTSRLGNGKPSRLDIDQKEIDBGEIFNAYQLTKTIPMGVYKIQG	241
	Sbjct: 181	FAFTPDKKTAIAEYTSR GRNG+ S+LD+D KELI+HGE+FN+Y L K+TIP GYK IQG	240
10	Query: 242	DAFVDKNNIAEVLNPSLESITISDXAPAIMSLKQVKLPDNLKIVIGELAFDMQIGGKINLP	301
	Sbjct: 241	DAFVDKNNIAEVLNPSLESITISDYAFAH++LKH+ LPDNLK IGLSAPDMQI GK LP	300
15	Query: 302	RHLIKLAERAFKSNRIOTVEFLGSKLKVIGRASPODNNLNVMFLDGLKIESEAFQGNP	361
	Sbjct: 301	R L++LAERAFKSN I+T+EF G+ LKVIIGRASPODNL+ +MLPDGLKIESEAFQGNP	360
20	Query: 362	GDEHYNNQVVLRTKRTQGNPHQLAETNTVYVNPDKSLMRATPDMDYTKMLBEDFTYQKNSVT	421
	Sbjct: 361	GD+HYNN+VVL T++G+NF LATENTIVNPKSLM+ +P++DYTKMLBEDFTYQKNSVT	420
25	Query: 422	GFSGKGLQKVRNNKNLEIFKQNGHITITIGDAFNFVDFQSKTLRKYDLESIKLPETIR	481
	Sbjct: 421	GFSGKGLQKVRNNKNLEIFKQNGH+TITIGDAFNFVDFQ+KTLRKYDLE+KLPETIR	480
30	Query: 482	KIGAFAPQSNILKSFEASDELSEIKGAPFNNRIOTGLDOKDLKIGDAAPHNNIYAIIV	541
	Sbjct: 481	KIGAFAPQSNILKSFEAS+DLESIKGAAPFNNRI TL+LKDKL+ IGDAAPHNNIYAIIV	540
35	Query: 542	LPESVQIGRSAPFRQWALHMFIGNKVTIGBNAFLSNKLESVNLSEKQKALTIYQAF	601
	Sbjct: 541	LPESVQELIGRSAPFRQGA +L+F+G+KVKT+GBNAFLSN+LE ++LSEKQK I VQAF	600
40	Query: 602	SDNALSEVVLFPNLQITIREBAFKNNHLEVKGSSTLSQITFNAPQNDQDKRPGKIVVVR	661
	Sbjct: 601	SDNAL EV+LP +L+TIREBAFK+NNH++ +S LS I FNA D NDG++F KVVV+	660
45	Query: 662	THNSHMLADGBRPIIDPDKLSTVLDLEKVLKIIIEGLDYSTLRQTTQTPQRMTTAGKA	721
	Sbjct: 661	THNSYALADGBRPIIDPDKLSTVLDLEKILKLIIEGLDYSTLRQTTQTPQRMTTAGKA	720
50	Query: 722	LLSKSNLRQSEKQKFLQRAQFFLGRVLDLKAIAKAERKALVTKKATNGHLLERSINKAVL	781
	Sbjct: 721	LLSKSNLRQSEKQKFLQRAQFFLGRVLDLKAIAKAERKALVTKKATNGHLLERSINKAVL	780
55	Query: 782	AYNNSAIKCANVKRLKEKLDLLTDLVBGKRGFLAQATMVQGVYLLKTPPLPPEYTYIGLVY	841
	Sbjct: 781	AYNNSAIKCANVKRLKEKLDLLT DLVBGKRGFLAQATMVQGVYLLKTPPLPPEYTYIGLVY	840
60	Query: 842	FDKSGKLIYALWSDTIGBQKDAYGNPILNVDRDNEGYHIAVATLADYBGLIKILIN	901
	Sbjct: 841	FDKSGKLIYALWSDTIGBQKDAYGNPILNVDRDNEGYHIAVATLADYBGLDIKILIN	900
65	Query: 902	SSLDKIKAIQIPLAKYHRIQIPQAIENNAABADRLLPKTPKGYLWVFNRYRQKVEKNL	961
	Sbjct: 901	S L ++ +IKH+P A YHR GIQPAI+NAABEA++LLPK	939
70	Query: 962	KPVDTKTPIPNKALPHEKVGDGRAGKHNNIAETNNSVAVTPIRSBQGLHESQSDVNLQP	1021
	Sbjct: 940	++++ + N++ ++S + ++ + LP+	977
75	Query: 1022	TSSKNPFYIILGYVSLCLLPVIVAGKKG	1051
	Sbjct: 978	T SK +F+Y IILGY S+ LL L+TA KK K	1007

SEQ ID 800 (GBS97) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 12; MW 113.4kDa).

GBS97-His was purified as shown in Figure 193, lane 6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 254

- 5 A DNA sequence (GBSx0269) was identified in *S. agalactiae* <SEQ ID 803> which encodes the amino acid sequence <SEQ ID 804>. This protein is predicted to be ribonucleoside-diphosphate reductase alpha chain (nrdE). Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4274 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB96160 GB:AE000050 ribonucleoside-diphosphate reductase alpha
 chain-MFN324 (new), 513 (Himmelreich et al., 1996)
 [Mycoplasma pneumoniae]
 20 Identities = 476/725 (65%), Positives = 586/725 (80%), Gaps = 20/725 (2%)
 Query: 2 TQSD--AYLSINAKTFDRGTGNYHTPSDKAEVQYMIHVEPTMTVPTSLIEKLDVLYS 59
 TQ D +Y+SLNA T+ F D AVE Y+ ERV+P T V E S E+LD+LV
 25 Sbjct: 12 TQEDLESYSLNATYTKVYG----DFQMDLHAVERAYIQERVKPTKTVFHSKZRLDPLVK 66
 Query: 60 NNYESDLLKQYNLEFICQIFERHAYAKKFAFLNFWGALKFPYNAYALKTEDNRYLYEHYED 119
 N+YI+ +++ Y+ E +I AYA +F +NFWG KFPYNAYALKT D +YLE+YED
 Sbjct: 67 NDYIDENINMTSFEQFSEITRKAYAYRFRYANFWGAKFYNYAYALKTFDGKYLYEHYED 126
 Query: 120 RVVNSALFLAAGDEKAYDLVDMDLANRFPQATPTFFLAGCKKRRGEYISCYLLRIEDNME 179
 RVVNS LFLA G+ A L+ ++ NRFPQATPTFFLAG+K+RGE+SCYLLRIEDNME
 30 Sbjct: 127 RVVNSVFLANGNYNKLAKLLKQIITNRFPQATPTFFLAGRKKRGEFVSCYLLRIEDNME 186
 Query: 180 SIGRAISTSLQSKRGGVVALCLTNLREPGAPIKGIQKATGIVPMKLLDESPSYANQL 239
 SI RAI+TWLQLSKR GQVAL LUN+RE GAPIK I+MQ+GI+PMKLLDESPSYANQL
 35 Sbjct: 187 SIGRAIYTTTLQSKRDGGVALLTNIRESGAPIKKIKENQSSGIIPINKLDESPSYANQL 246
 Query: 240 GQRQAGAVYLHAHPEVLVFLDTKKNDADEKIRIKSLGLVLPDITFELAKANKDMAL 299
 GQRQAGAVYLHAHHP+V+ FLATKKNDADEKIRIKSLGLVLPDITF LAK N++HAL
 40 Sbjct: 247 GQRQAGAVYLHAHHPDVMQFLDTKKNDADEKIRIKSLGLVLPDITFELAKANKDMAL 306
 Query: 300 FSPYDIERYVGKPMYSDISITETEEYETLLANADIKTFISARKLFOTIAELHFESSYPIYLF 359
 FSPYD+ YQK+SDIS+TE Y LLAN I+KTFI+ARK FOT+IELHFESSYPIYLF
 45 Sbjct: 307 FSPYDVIYEGKPLSDISVTEYETELLANQRIKTFINARKFPOTIAELHFESSYPIYLF 366
 Query: 360 EDTVNAKNPHKKBRIWMNLCRSIAQVNTASQPSSEDITLVKQWHDVCCNLGNSINIAARM 419
 +DTVN +N H RIVMSNLCSEI Q +T S+F DL F KVG+D+ CNLGS+NIA+AM
 50 Sbjct: 367 DDTVNRNDAH--FNRIWMNLCSEIVQSPSTPEPHDLAKFKVGNDISCNLSINIAKAM 424
 Query: 420 DQADFEKLIANSIRALDRVSRKTDLDSEAPSIIKKNDAANHAVGLGAMNHLGFLATNIIYY 479
 + +F +L+ +I +LD VSR S+L+APS I+KGN+ NHA+GLGAMNHLGFLATN IYY
 55 Sbjct: 425 ESGPFPSELVLAIESLDLVSKVSNLEFADSIQKGNSENHALGLGAMNHLGFLATNIYY 484
 Query: 480 DQGEAIDFTDCFFYAMAYYAKSNHSLAKEKTFPESSESSYADGSYFYQY--TEQNF-E 536
 +S EADFT+ FFY +AY+AFKAS+ LA EKG F+ F +ADGSYF +Y E +F
 60 Sbjct: 485 NSPEAIDFTINIFTTVAYHAFKASLEALEKGFKNPENTKFDAGSYFDKYIKVSDPFWT 544
 Query: 537 PKTRVKNLLAEYGLITLPSQEDWREKLVQSIKEISLANAHILAAPTGSISYLSSTCPSLQ 596
 PKT+RVK L +Y +P+ +E+W+L +I++ GLAN+HLA+APTGSISYLSSTCPSLQ
 60 Sbjct: 545 PKTERVKALFQKQYELPTRENNKELALNQLKGNLANSHLILAAPTGSISYLSSTCPSLQ 604

Query: 597 PVVSPVEVRKEGALGRVYVPAYKIDADNYVYKKGAYEVGSEAIINIAAAQKHIDQALS 656
 PVVSPVEVRKEG LGR+YVAY+++ D+Y +YK GAYE+G E IINIAAAQ+H+DQALS
 Sbjct: 605 PVVSPVEVRKEGLRIYVAYQLNKDSYFPYKDGAYELGPPEIINIAAAQGVHQAIS 664

5 Query: 657 LTLFMTDQATTTRDLNKAYIAFKQKCSIIYVVRQDILEGSSYDDMLDFTSSDLEDC 716
 LTLFMTD+ATTTRDLNKAYI AFK+ C+SIYVVRVQ++LE SE + + ++ C
 Sbjct: 665 LTLFMTDQATTTRDLNKAYIYAFKKGCSIIYVVRVQRVELEDSIH-----TIQQQC 716

10 Query: 717 QSQMI 721
 ++C+I
 Sbjct: 717 EACVI 721

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 805> which encodes the amino acid sequence <SEQ ID 806>. Analysis of this protein sequence reveals the following:

Possible site: 47
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1843 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

25 >GP:AA02625 GB:AF054892 surface antigen BspA [Bacteroides forsythus]
 Identities = 124/451 (27%), Positives = 202/451 (44%), Gaps = 65/451 (14%)

Query: 221 PMSYLLKVTIPYGYKHIGQDAFVDNKNIAEVLNPLESLTISDYAFALHA-LWQIDLDPN 279

30 F+ L VT+P IG AF + + +P S+ TI ++AF + LK I LP++
 Sbjct: 87 FSDCALVTSLPNSLTAIGDAHAFKGCGLTSITIPNSVTITIGWAFKGCGLKSLITLPS 146

Query: 280 LKAIGELAFDDNQITOKLSLPRQLNRAERA-FKSNHIKTIEFRGNSLKVIGESQD-N 337
 L AIG+ A +++P + + E A F + + +I F NSL IGE++F

35 Sbjct: 147 LTAIGCALSGCTGLTSITIPNSVTITIGWAFKGCGLTSITF-PNSLTAIGESAFYQCG 205

Query: 338 DLSGLMLPDGLEKIESEAFNGPDGDHYNRVVLMTKSGKIPSGLATENTYVNPDKSLMQ 397
 L+ + LPD L I AF G G KS P+ L T +S +

40 Sbjct: 206 ALTSITLPAALTIGESAFKGCSCG-----LKSITFPNSLITIG-----BSAFY 248

40 Query: 398 ESPEIDYTWLEEDFTYQKNSVTFGNSNGLQKVRKNKLEIPKQNGVTITEIGDAPRN 457
 + + + T +++ G S GL K++ P ++T IG++AF N

Sbjct: 249 DCGALTSITLPAALTIGESAFKGCSCG-----LKSITFPNSLITIG-----BSAFY 295

45 Query: 458 VDFQNKTLRKYLDEEVKLPSTIRKIGAPAFQS--NNLKSFEASDLSEIKGAPMNRRET 516
 L + +P+++ IG AF + LKS D L I+E AF N + T

Sbjct: 296 CG-----SLTSITIPNSVTITIGESAFKGCGLKSLITLPGDLTTEIRAFYNGVLT 346

Query: 517 -LELKDILVTIGDAPH-INHIYAVLPESVQIEGRSAFRQNGANNLIPMGSKVKTIGEM 574
 + + + + TIG++AF+ + + I LP+ + I AF GA I + V T+GE

50 Sbjct: 347 SITIPNSVATIGESAFYQCGSLKSLITLPGDLTTEWGAFFNGALTSITIPNSVTIGES 406

Query: 575 AFLS-NRLEHLDLSEQKQLTRIPVQAPSDNALKEVIL--PASLKTIREAPKKNHLEQL 631
 AF L+ + ++ + +I F + L + L PA KT+ E K+ K+

55 Sbjct: 407 AFYGCALGDVTVAWDTPI-DIQEDVSRRLTSLGRLHVPAGKKTVE--ADWVKV-- 460

Query: 632 VASALSHIAFNALDDND-GDEQFINKVVVKT 661
 FN ++D+D G Q++ KT

Sbjct: 461 -----FNIVEDDDGGGLQNNYDAATKT 482

60

An alignment of the GAS and GBS proteins is shown below:

Identities = 534/726 (73%), Positives = 614/726 (84%), Gaps = 5/726 (0%)

Query: 1 MTQSDA-YLSLNAKTRFRDRTGNVHTSDKEAVEQYMIHVEPVMVMTSLIEKLDYLVLS 59

-320-

```

      M+Q++A YLSINA TRF+   G+YHP SOKRAV +Y+ HNV PN M P SL +XL YL++
      Sbjct: 1 MSQTNASYLSINALTRFKKPGQSYHFDSDOKRAVRRYLREHVSPNQMAPNSLEDKLAYLIN 60

5   Query: 60 NNYTESDLLKQYNLEPIQCFEHAAYAKKPAFLNFMGALKPTFHAVALKTHRYLYEHVED 119
      YTB + Y + I + F +RY + + FLN MGA+KFT +YALKT D + YLE +ED
      Sbjct: 61 BGYEQALFDAYPNDLIKEAFHAYQQSTRFLNLMGAMKFTYQSYALKTKLQKQYLETSPD 120

10  Query: 120 RVVNNALFLARGDEKAYDLVDOMANRFPQATPTFFLNAGKKRGEYISCYLLRIEDME 179
      R VNNALFLA GD+ +D+D +L RFQPATPTFFLNAGKKRGEYISCYLLRIEDME
      Sbjct: 121 RAVNNALFLADGQDTFVFVDLAILHRRFPQATPTFFLNAGKKRGEYISCYLLRVEDME 180

15  Query: 180 SISRAISTSLQSKRGQGVALLC/TNLRERGAQIKGIGNQATGIVPMVKLLSDSFSYANQL 239
      SISRAISTSLQSKRGQGVALLC/TNLR E GAPIKGI +NQATGIVPMVKLLSDSFSYANQL
      Sbjct: 181 SISRAISTSLQSKRGQGVALLC/TNLR EIGAPIKGIENQATGIVPMVKLLSDSFSYANQL 240

20  Query: 240 GQRQAGAVYLHAHHPEVLFTLDTKRENADEKIRIKSLSLGLVLPDITFELAKANKDAL 299
      GQRQAGAVYLHAHHPEVLFTLDTKRENADEKIRIKSL+LGLVLPDITF+LAK NCKDAL
      Sbjct: 241 GQRQAGAVYLHAHHPEVLFTLDTKRENADEKIRIKSLALGLVLPDITFQAKENKDAL 300

25  Query: 300 FSPFDIERVYVKPMDSISITEYETLLANADIRKTFISARKLPQTIAELHFSGYPTLFF 359
      FSPFDI+R YGK MSDISITEY+ LLAN I+KT+ISARK FQ IABLHFSGYPT+LF
      Sbjct: 301 FSPFDIKRAYGKMSDISITEYIDKLLANPAIKKYIISARKFFQLIABLHFSGYPTLFF 360

30  Query: 360 EDTVNAKNPHKKGRIVMSNLCSEIAQVNTASQPSBOLDTPTKVGHEDVCCSLGSINIAARM 419
      +DTVN +NPH K+GRIVMSNLCSEIAQV+T S F EDL+F +G D+CCSLGSINIA+AM
      Sbjct: 361 DDTVNKNPNHAKKGRIVMSNLCSEIAQVSTSTFKEDLSFETIGEDICNLSGINIAARM 420

35  Query: 420 DQADPEKLIANSIRALDRVSRSDLSAPSIIKKGNANHAIVGLGAMNHLGFLATHIYY 479
      A FE+LI SIRALDRVSR SDL+ APS+ + GNAANHAIVGLGAMNHLGFLATHIYY
      Sbjct: 421 ADAPHEPQLITTSIRALDRVSRSDLNCAFSVETGNANHAIVGLGAMNHLGFLATHIYY 480

40  Query: 480 DSGEALDFDCFFMAYAYAFKASHNLAKSKGTFEGPSESSYADGSYPYQTBQNFPEKT 539
      D+ +EA+DPTD FF+MAYAYAFKAS LAKEKQ F GPS S+Y+DG+YF +Y +AA +P+T
      Sbjct: 481 DTKEAVDPTDLFFMAYAYAFKASQLAKEKQFAGFSLSTYSOGTYPAKYLQEDAKFQT 540

45  Query: 540 QRQVGLLAIEGLTLPSCENRHLVQSIKEIGLANAHLAVAPTQSIYLSCTPSIQPVV 599
      +V LL +YQ TLP+ DN+ LV IK+ GLANAHLAVAPTQSIYLSCTPSIQPVV
      Sbjct: 541 AKVATLLQDYGPTLPFLVADWQALVADIKQPLANAHLAVAPTQSIYLSCTPSIQPVV 600

50  Query: 600 SPFEVRKEGALGRVYVPAYIKDANVYVYKKGAYEVGSEAIINIAAAQKHIDQAISLTL 659
      +PFEVRKEG+LGR+YVPAY-ID NY YY++GAYEVG +AII++ AAAQKH+DQAISLTL
      Sbjct: 601 APFEVRKEGSLGRIVYVPAYIQDQANYAYERGAYEVGPKAIIIDVVAQAQKHVDQAISLTL 660

60  Query: 660 PMTDQATTDLNKAIVQAPKQCASIIYVVRVQDILEGSEYDQ---MLDDPSESELED 715
      PMTDQATTDLN++YICAPKQ CASIIYVVRVQD+L GSE YD+ + +
      Sbjct: 661 PMTDQATTDLNRSYIQAPKQCASIIYVVRVQDVLAGESEYQDEDSLVTPAPGASDETITE 720

      Query: 716 CQSCMI 721
      CQSCMI
      Sbjct: 721 CQSCMI 726

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 255

- 55 A DNA sequence (GBSx0270) was identified in *S. agalactiae* <SEQ ID 807> which encodes the amino acid sequence <SEQ ID 808>. This protein is predicted to be nrdI protein (nrdI). Analysis of this protein sequence reveals the following:

```

Possible site: 54
>>> Seems to have no N-terminal signal sequence

```

60

```

----- Final Results -----

```

```

bacterial cytoplasm --- Certainty=0.2952 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

```

-321-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

5 >GP:AACT1451 GB:U39702 nrdI protein (nrdI) [Mycoplasma genitalium]
 Identities = 77/127 (60%), Positives = 104/127 (81%), Gaps = 1/127 (0%)

Query: 7 VVYFSKSNNTHRFVQKLACSNQRIPSD-GSSILVTEYILIVPTYAGGGDDTKGAVPKQ 65
 +VYFSS SNNTHRF+KL ++RIP D SI V+ +Y+LI PTY+GGG+ +GAVPKQ
 10 Sbjct: 22 IVYFSSIGNNTHRFIEKLGQFQHKRIPVDITQSTVYNEVYLICTYSGGGNVEGAVPKQ 81

Query: 66 VVQFLNVRQNRHCCQGVISGNTNFGDTYALAGPIIARKLNVPLLHQFELLGTQEDVTRV 125
 V+QFLN + NRE C+GVI+SGNTNFGDT+ +AG +I++KLNVELL+QFELLGT+ DV +
 15 Sbjct: 82 VIQFLNKHNRHRELGRGVIASGNTNFGDTFCNGTVISKLNVELLHQFELLGTKNDEVT 141

Query: 126 KELLCOF 132
 ++++ ?
 Sbjct: 142 QKLIANF 148

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 809> which encodes the amino acid
 sequence <SEQ ID 810>. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 25 bacterial cytoplasm --- Certainty=0.0089 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

30 Identities = 84/125 (67%), Positives = 100/125 (79%)

Query: 7 VVYFSKSNNTHRFVQKLACSNQRIPSDGSILVTEYILIVPTYAGGGDDTKGAVPKQV 66
 +VYFSSKNNTHRFVQKL GRIP D + V+ Y+LIVPTYA GG D KGAV KQV
 35 Sbjct: 6 IVYFSSKNNTHRFVQKLGILPAQRIPFVDRNRELVSTHYLLIVPTYAGGSDAKGAVSKQV 65

Query: 67 VVQFLNVRQNRHCCQGVISGNTNFGDTYALAGPIIARKLNVPLLHQFELLGTQEDVTRV 126
 ++FLN NR+HC+GVISGNTNFGDT+A+AGPII++KL VPLLHQFELLGT DV +V+
 35 Sbjct: 66 IRLFLNPNRKHCKGQGVISGNTNFGDTFALAGPIISQKLQVPLLHQFELLGTATDVTRVQ 125

40 Query: 127 ELLCQ 131
 + +
 Sbjct: 126 AIFAR 130

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 45 vaccines or diagnostics.

Example 256

A DNA sequence (GBSx0271) was identified in *S.agalactiae* <SEQ ID 811> which encodes the amino acid
 sequence <SEQ ID 812>. This protein is predicted to be ribonucleoside-diphosphate reductase beta chain
 (nrdF). Analysis of this protein sequence reveals the following:

50 Possible site: 27
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 55 bacterial cytoplasm --- Certainty=0.3889 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

-322-

>GP:AB56162 GB:AE000050 ribonucleoside-diphosphate reductase beta chain [Mycoplasma pneumoniae]
Identities = 261/335 (77%), Positives = 301/335 (88%)

5 Query: 2 QSYVDRSQPLDYALSEKAFPMRSVWNNKLNDDKLEVWNRVTQNFWLPEKIPVSNLMS 61
+ Y+ S SPL+YA + +RSVWNN ++D+KDELVWNR+TQNFWLPEKIPVSN+D S
Sbjct: 5 KKYFLSVSPLEYAQKPKQGNLRVWNNKLNDDKLEVWNRVTQNFWLPEKIPVSN+D S 64

10 Query: 62 WRTLDADMQQLITRTFTGLTLLDSVQATVGDIAQIKHSQTDHEQVIYANFAPMVAIHARS 121
W+ L +MQ LIT+FTGLTLLD++QAT+GDI QI ++ TDHEQVIYANFAPMVAIHARS 124
Sbjct: 65 WQLSKSEWQQLITRTFTGLTLLDTIQATGDIKQIDYALTDHEQVIYANFAPMVAIHARS 124

Query: 122 YGTIFSTLCTSQIEBAHEWVVDIESLQARSRLIPFTYTDGDDPLKSKVAAAMMPGFLLYG 181
YGTIFSTLCT+QI BAHEWVV TESLQ R++ LIP+ITG DPLKSKVAAA+MPGFLLYG 184
Sbjct: 125 YGTIFSTLCTSEQITRAHEWVVDIESLQKRAKALIPYITGDDPLKSKVAAALMPGFLLYG 184

Query: 182 GFYLPFYLSARGKLNTSDIIRILIRDKVIHNNYSGYKQKVAKLVSQKAEKMTFVFD 241
GFYLPFYLS+R +LNTSDIIRILIRDKVIHNNYSGYK+Q+RV R+S EKQAEK FVFD 244
Sbjct: 185 GFYLPFYLSRKLNTSDIIRILIRDKVIHNNYSGYKQKVEKMSKQAEKMTFVFD 244

20 Query: 242 LLYQLIDLEKAYLRELYVGDFLAEDAIRFSYINAGKFLQNLGYDSPFTTESTRISPEVFA 301
L+Y+LI+LEKAYL ELY+GF + EDAI+FSYINAGKFLQNLGYDSPFTTESTR I PE+FA 304
Sbjct: 245 LMYELILEKAYLRELYVGFVEDAKFSYINAGKFLQNLGYDSPFTTESTRIPKIFA 304

25 Query: 302 QLSARADENHDFPSGNGSSYIMGITETLDEWDF 336
QLSARADENHDFPSGNGSSY+MGI+ET D+DW+P
Sbjct: 305 QLSARADENHDFPSGNGSSYVMGITETEDKDWDF 339

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 813> which encodes the amino acid sequence <SEQ ID 814>. Analysis of this protein sequence reveals the following:

Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
35 bacterial cytoplasm --- Certainty=0.3779 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

40 Identities = 292/335 (87%), Positives = 318/335 (94%)

Query: 2 QSYVDRSQPLDYALSEKAFPMRSVWNNKLNDDKLEVWNRVTQNFWLPEKIPVSNLMS 61
Q YY+RSQSP++YALSE +RS+NNN LNDDKLEVWNRVTQNFWLPEK+PVSNLMS 62
Sbjct: 3 QHYVRSQSPIEYALSEFTQQLRSINWNNKLNDDKLEVWNRVTQNFWLPEKIPVSNLMS 62

45 Query: 62 WRTLDADMQQLITRTFTGLTLLDSVQATVGDIAQIKHSQTDHEQVIYANFAPMVAIHARS 121
WR+L DMQQLITRT+FTGLTLLD+VQATVGD+AOI+HSCQTDHEQVIY NFAPMVAIHARS 124
Sbjct: 63 WRSLOEDMQQLITRTFTGLTLLDVQATVGDVQAQIHSQTDHEQVIYINFAPMVAIHARS 122

50 Query: 122 YGTIFSTLCTSQIEBAHEWVVDIESLQARSRLIPFTYTDGDDPLKSKVAAAMMPGFLLYG 181
YGTIFSTLC+S+QIEBAHEWVV T+SLQ R+R+LIP+ITGDDPLKSKVAAAMMPGFLLYG 184
Sbjct: 123 YGTIFSTLCTSEQIEBAHEWVVDIESLQARARVLIPTTYTDGDDPLKSKVAAAMMPGFLLYG 182

55 Query: 182 GFYLPFYLSARGKLNTSDIIRILIRDKVIHNNYSGYKQKVAKLVSQKAEKMTFVFD 241
GFYLPFYLSARGK+PNTSDIIRILIRDKVIHNNYSGYKQKVA+LS EKQAEK FVFD 244
Sbjct: 183 GFYLPFYLSARGKMPNTSDIIRILIRDKVIHNNYSGYKQKVARLSPEKAEKMTFVFD 242

Query: 242 LLYQLIDLEKAYLRELYVGDFLAEDAIRFSYINAGKFLQNLGYDSPFTTESTRISPEVFA 301
LLY+LIDLEKAYL ELY GFLAEDAIRFSYINAGKFLQNLGY+SPFT+EKTR+SPVFA 304
Sbjct: 243 LLYELIDLEKAYLRELYVGFVEDAKFSYINAGKFLQNLGYDSPFTTESTRISPEVFA 302

Query: 302 QLSARADENHDFPSGNGSSYIMGITETLDEWDF 336
QLSARADENHDFPSGNGSSY+MGI+ET D+DWDF
Sbjct: 303 QLSARADENHDFPSGNGSSYVMGITETIDDEWDF 337

-323-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 257

A DNA sequence (GBSx0272) was identified in *S.galactiae* <SEQ ID 815> which encodes the amino acid sequence <SEQ ID 816>. This protein is predicted to be rhamnosyltransferase. Analysis of this protein sequence reveals the following:

Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1741 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9583> which encodes amino acid sequence <SEQ ID 9584> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAAS2090 GB:AB010970 rhamnosyltransferase [Streptococcus mutans]
Identities = 104/309 (33%), Positives = 173/309 (55%), gaps = 21/309 (6%)

Query: 11 QINICLATYNGOKYLRQQLDSIIQQGYIDWICLRDDGSTDDTVAIKEVNRDSRFIF 70
++NI ++TYNGQ++ QQ+ SI +Q +.W LIRDDGS+D T II ++ D+R FI
Sbjct: 2 KVINLMSTYNGQEPFAQQIQSTQKOTFEMNNLLIRDDGSDGTPKIADFAKSDARIFI 61

Query: 71 NSNDRKLGSHRSFYELVNYKADYFVSDQDDWKENRIERYLEEAKFNQSLPLLVYS 130
N++ G ++FY L+ Y+KAD+Y FSDQDDW +LE L EK N ++PL+VY+
Sbjct: 62 NADKRENFVINKNFYITLLKYEKADYFVSDQDDWLPQKLELTLASVEKENNQIPLAVYT 121

Query: 131 NWTSDVKELTVL-----KSHNPATVIOEQIAFNQINGMVMNNHCLAKWE--YRQIG 181
+ T VD L VL + H+ T + E+ N + G +M+NH LAK W+ Y +
Sbjct: 122 DLTVDVDRDLQVLDHSMIKTQSHANTSLLEELTENTVTGSTMVNNHCLAKQWQCYDDLI 181

Query: 162 AHDSTVGTLAYAVGNVAYISDSTVLMRRQ---VGAES---LNNYGRQYG-VATFWQMI 232
HD Y+ LA ++G +Y+ ++T L+R+ +GA + L N+ R + V +W ++
Sbjct: 162 MHDWYLAALLAASLGKLIYLDFTTLLYRQHESNVLGARTWSKRLKNNLPHRLVKKYMWLV 241

Query: 233 NTSFDRASLIFAQVSDKMSLERKLFFSRFIELKNANIMRRIYLLSKLERKSLKETVAN 292
+S +AS + + + K ++ L + + + RI L + + T
Sbjct: 242 TSSQQQASHL---LELDLPANKAIIIRAYVTLLDQSFNRIKWLKQYGFKNRAPHFTVP 298

Query: 293 TILLITGYG 301
L++T +G
Sbjct: 299 KTLIITKYG 307

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 258

A DNA sequence (GBSx0273) was identified in *S.galactiae* <SEQ ID 819> which encodes the amino acid sequence <SEQ ID 820>. Analysis of this protein sequence reveals the following:

Possible site: 35
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -4.19 Transmembrane 1213 -1229 (1211 -1230)

----- Final Results -----

bacterial membrane --- Certainty=0.2678 (Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9581> which encodes amino acid sequence <SEQ ID 9582> was also identified.

There is also homology to SEQ ID 822.

A related GBS gene <SEQ ID 8525> and protein <SEQ ID 8526> were also identified. Analysis of this protein sequence reveals the following:

10 Lipop: Possible site: -1 Crend: 7
 SRCFLG: 0
 MoG: Length of UR: 3
 Peak Value of UR: 2.28
 Net Charge of CR: 4
 15 MoG: Discrim Score: 1.29
 GVH: Signal Score (-7.5): 2.84
 Possible site: 30
 >>> Seems to have a cleavable N-term signal seq.
 Amino Acid Composition: calculated from 31
 ALOM program count: 0 value: 1.16 threshold: 0.0
 20 PERIPHERAL Likelihood = 1.16 344
 modified ALOM score: -0.73
 *** Reasoning Step: 3
 25 ----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
 30 LPXTG motif: 1197-1201

SEQ ID 8526 (GBS147) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 4; MW 132kDa).

The GBS147-His fusion product was purified (Figure 200, lane 5) and used to immunise mice. The resulting antiserum was used for FACS (Figure 286), which confirmed that the protein is immunoreactive on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 259

40 A DNA sequence (GBSx0274) was identified in *S.galactiae* <SEQ ID 823> which encodes the amino acid sequence <SEQ ID 824>. This protein is predicted to be Acetyltransferase (GNAT) family. Analysis of this protein sequence reveals the following:

Possible site: 57
 >>> Seems to have no N-terminal signal sequence
 45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2781(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 50

The protein has homology with the following sequences in the GENPEPT database:

>GP:RAG03505 GB:AB004449 conserved hypothetical protein [Pseudomonas aeruginosa]

-325-

Identities = 66/143 (46%), Positives = 94/143 (65%), Gaps = 5/143 (3%)

Query: 2 WNVKTFDNLTFHFLFOYKLRVSVFVVDQCPYQEVDDHLLI--CLAGMHWVDGQLANY 59
 W K +LT EL+ + +LR VPFVEQ CPYQEV D L+ H M W DQQL AY
 5 Sbjct: 5 WICKHHADLTAKELYALLQKRTVFVVDQKCPYQEVGDLVGDTHHLMARWDQQLAYL 64

Query: 60 RLIP---EDDKVHLGRVIVNPDPRKKGGLGNOLVEYALKPSRANVFNKPIYAQAQAYLQDF 116
 RL+ + +V +GRV+ + R +GLG+QL+E A++ +E + + P+Y AQA+LQ +
 10 Sbjct: 65 RLIDPVZHEGQVVICRVSSSARGGGLGHQMERALQAAERLWLDTPVYLSAQAHQAY 124

Query: 117 YGSFGFQFVSDTYLEWIPHLDM 139
 Y +GF V++YLED+IPH+ M
 10 Sbjct: 125 YGRYGFVAVTEVLEDDIPHIGM 147

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 260

A DNA sequence (GBSx0275) was identified in *S.agalactiae* <SEQ ID 825> which encodes the amino acid
 20 sequence <SEQ ID 826>. Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2010 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

30 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 261

A DNA sequence (GBSx0276) was identified in *S.agalactiae* <SEQ ID 827> which encodes the amino acid
 35 sequence <SEQ ID 828>. Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2935 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

45 >GP:CAB12631 GB:Z99108 similar to RNA methyltransferase [Bacillus subtilis]
 Identities = 217/448 (48%), Positives = 298/448 (66%), Gaps = 4/448 (0%)

Query: 7 QRIPLKIKRMGNGINGEGIPYKTLIPVPGALKEGEVFCQISSVRNPNPAKLLKIKGSK 66
 Q PL IKR+GINGS+G++KK ++FVPGAL GEEV Q + V+ F+E ++ KI K S+
 50 Sbjct: 16 QTFFLTIRKLINGSGBOVGYPKKVVFVPGALFGEVVQATKVKQPKSGRKKIKRSGE 75

Query: 67 NRVEPPCSYIKYKCGSQIMELQYDKQLGPKTDVIRQALMKFKPKGEYNYEIRKTIQSGEP 126
 +RV PPC +Y++CGSQ+ EL Y +QL K D++ Q+L + EN EI++TIGM P

-326-

Sbjct: 76 HRVAPPCCPVYEQCGCQQLHLAYSQQLREKRDIVIQSLRRHFKPKVRNMKIKSTIGMDNP 135
 Query: 127 EHYRAKLQFQV-RSPGNGVKGALYAQGTHERLIDIKDCLVQDSLTOGMNIRVAELLGKYKL 185
 +YR K QPC+ RS G++ AGLY +H ++ IKDC+VQ T + V +L ++ +
 5 Sbjct: 136 WYRNKQSQFQIGRSQSGSIIAGLYGLSDHIVPIKDCIVQHPATNKTTIGVRIILEDPIV 195
 Query: 186 FIYNERKIAG-VRTVMIRRAQASGEVQLIFITSKRL--DFDDVVIELVREPELKTAVN 242
 +YNERK G VRT++ R +GEVQ++ +TK +++V + + PE+K++ N
 10 Sbjct: 196 SVTNERTRKGDVRTIVTRVGPEISGVQVVLVTAKETLPHKEIVKAIQKRLPEVKSIIQN 255
 Query: 243 INASKTSIDIYQITEVINGQBSINEEVLVDYGFSLSPRAFYQLNPKQTQILYSEAVKALDV 302
 +N +KTS I+G+ T+ + G+ I E + D F L S RAP+QLNP+QT LY E KA + +
 Sbjct: 256 VNGAKTSVIFGKTKQLAGKTVIQEVLGDVFSKLGRAPFQLNPEQTVKLYDEVKGAQL 315
 Query: 303 KEDDLIDAYCGVGTIGLAPAGKVKSVRGMDIIPBAIQDAKENALYMGFTNTHYEAGKAE 362
 + ++DAYCGVGTIG+ A K VRGMD+I R+I DAK+NA G N Y G AE
 15 Sbjct: 316 TSGEKVVDAYCGVGTIGM+VADGAKEVRGMDVIBESIDDAKNAKKGMANATYVTGTAE 375
 Query: 363 DIIPKYSSEGFANALIVDPPTGLDCKLNTILMPPKPMVYVSCNISTLARDLVLTAK 422
 +P+N ESFR + +IVDPPTG D L+TI K+ P+ VYVSCN SLA+DL TL+K
 20 Sbjct: 376 HWLAKTKKEGFRDPPVIVDPPTGCDSTFLTITKVKPKRFTVYSCNISTLANDQLTSLK 435
 Query: 423 VTRVHYIQSDVMFPHARTAEAVKLRK 450
 Y V YIQ VDMFP TA EAV +L K
 25 Sbjct: 436 DYKVDYIQSDVMFPQTAHVEAVKRLK 463

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 829> which encodes the amino acid sequence <SEQ ID 830>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence
 30
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2980 (Affirmative) < succ>
 bacterial membranes --- Certainty=0.0000 (Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 327/450 (72%), Positives = 397/450 (87%)

40 Query: 1 MNVVLKQRIPLKIKRMGINGEGIFYKKTILFVPGALKGEVFCQISSVRNFAEAKLLK 60
 N V +KQ+IPLKIKRMGINGEGIFY+KTLFVPGALKGE++PQI++V+RNFARAKLL
 Sbjct: 1 MNVVKQKIPLKIKRMGINGEGIFYKKTILFVPGALKGEDIFCQITAVRNFAEAKLLT 60
 Query: 61 INKSKIRVPEPCSIYKECGCQIMHLQYDKLEFKTIDVIRQALMKFKPKPGYENVEIRKT 120
 +NK SKIRV+P CS+Y+ CGCQIMHL Y KRL+PK DVIRQAL KFKP GVE +EIR T
 45 Sbjct: 61 VNGSKIRVVKPACSVYETCGCQIMHLAYPKQLDPRDDVIRQALMKFKPVTGQFEIRPT 120
 Query: 121 IGMSEPHYRAKLQFOVRSFGNGVKGALYAQGTHERLIDIKDCLVQDSLTOGMNIRVAELL 180
 +GM +P+HYRAKLQF+RSPG VKAGL++QG+HRL+ I +CLVQD LTO++IN++ +L+
 50 Sbjct: 121 LGMKKPHYRAKLQFQLRSPGTVKGLPSQSGEHLRIPDNCLVQDQITQINDIKITQLV 180
 Query: 181 GKVKLPYNERKIAGVRTVMIRRAQASGEVQLIFITSKRLDFDDVVIELVREPELKTAVN 240
 KVKLPYNERKIAG+RT+M+R+AQAS +VQ+I ++SK + + + EL + PP++KTV
 55 Sbjct: 181 DKVKLPYNERKIAGRTIMVRKAQASDQVQIIVVSGSEKVLNFIKELTAFQPVKTV 240
 Query: 241 VNINASKTSIDIYQITEVINGQBSINEEVLVDYGFSLSPRAFYQLNPKQTQILYSEAVKAL 300
 +N N SK+S+IYG TE++WQGE+I+HEVLVDYGF+LSRAPYQLNP+QT++LY E VKAL
 Sbjct: 241 LNENRSKSSIYQDETEILWQQAIEHEVLVDYGFALSPRAFYQLNPQTEVLYGEVVKAL 300
 Query: 301 DVKEDDLIDAYCGVGTIGLAPAGKVKSVRGMDIIPBAIQDAKENALYMGFTNTHYEAGK 360
 DV D +IDAYCGV+IG A PAGKVKSVRGMDIIPBAIQDAKENALYMGFN N NGF N +YEAGK
 60 Sbjct: 301 DVGSKDHIIDAYCGVSGIGAPAGKVKSVRGMDIIPBAIQDAKENALYMGFNATYAEAGK 360
 Query: 361 AEDIIIPKYSSEGFANALIVDPPTGLDCKLNTILMPPKPMVYVSCNISTLARDLVLT 420
 AEDII +WT +G+RA+AI+VDPPTGLDCKL TIL P++MVVYVSCNISTLARDLV L
 65

-327-

Sbjct: 361 AEDIISKWYKQGYRADAVIVDPPTGLDOKLLEKTLIHYQPKQWVYVSNSTLTDLVQL 420

Query: 421 TKVYHVHYIQSDVMFPHETARTEAVVKLQK 450

5 Sbjct: 421 TKVYDVHYIQSDVMFPHETARTEAVVKLQK 450

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 262

- 10 A DNA sequence (GBSx0277) was identified in *S. galactiae* <SEQ ID 831> which encodes the amino acid sequence <SEQ ID 832>. Analysis of this protein sequence reveals the following:

Possible site: 45
>>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3505 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 20 The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04643 GB:AP001510 unknown conserved protein in B. subtilis
[Bacillus halodurans]
Identities = 74/263 (28%), Positives = 141/263 (53%), Gaps = 9/263 (3%)

- 25 Query: 3 ITKIEKIGGR---LYTLEL-DNTEINLY---ITEDTIVHFMISKGMINAEKLENIKKFAGL 55
IT+IE +KR Y + + N +++Y + E ++ L KG+ I+AE+++ I ++
Sbjct: 4 ITKIEVQKGNRRYNTIPIHQNGQDVYAFSVDQVLTIKQGLRKGLDIDABGMKQILYEDV 63

- 30 Query: 56 SYGNLGLYIISFKQRTSEKVIKYLQCHDIDSIIPIQIIDLKSEMWINDKNVVSQFIQQ 115
NL L+Y+S++ R+ EV YL++ D + II ++ L + + ++D + ++FIQ
Sbjct: 64 QKTFFNLALHYLSYMRSVHEVRTYLKKKDREEFIIHVLHRLTEQRLDDHAFARAFIT 123

- 35 Query: 116 NLNTGDKGPFYVIKQKLLQGIKSIIESELQAINPQDLASKISQKLYKXYQKLPALQAL- 174
T KGP +KQ+L +KG+ K IE L +++++ ++ L R+ +L
Sbjct: 124 KRATTSKGLPKLQELAEKGVSEKNTIEGALTTPSYEEQVEQVAVLERKQGRFTFGSSLA 183

- 40 Query: 175 -KKKMQSLTTKQPDYQIVHTVIGNLEIKDQLEEDLYLKDFQYKLSKHIDQYELK 233
K KL + L KG+ ++ I++++E R + + +K +K + K +EL+
Sbjct: 184 WKQICSRQLLAKGYTSPVTEAFADVPPIKQEEEEWEALKAFGEKMKRYAGKKTQWELQ 243

- 45 Query: 234 QRILNALMRKGYQYEDIKSALRE 256
Q++ AL RKQ+ E + L +
Sbjct: 244 QKVQALYKRGFSLEMIERYLAD 266

- 45 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 833> which encodes the amino acid sequence <SEQ ID 834>. Analysis of this protein sequence reveals the following:

Possible site: 58
>>> Seems to have no N-terminal signal sequence

- 50 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2388 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 55 An alignment of the GAS and GBS proteins is shown below:

Identities = 146/258 (56%), Positives = 190/258 (73%)

Query: 1 MKITKIEKKRKYLTLELNTENLYITTEDTIVHFMISKGMINAEKLENIKKFAGLSYGN 60
MKITKIEKKRKY +ELDN E+LY+TSDTIV FMISK +++ ++LE+K FAGLSYGN

-328-

Sbjct: 1 MKITKIEKKRLLYLIELDNDESLVYTEDITVRFPLMSKDKVLNDQLEDMKHFAQLSYGRN 60
 Query: 61 LGLYTIISFKQRTKEVVIKYLQOHIDISKLIFQIIMLKSENWINDKNVVSPTQQNNTG 120
 L LY++SF+QR+ K+V YL++H+I+ II II L+ E WI+D ++I+QN G
 5 Sbjct: 61 DALYPLSFQQRSNKQVADYLRKHIEIRHTIADITTYLQBSQWIDDTKLAITYIRQNQLAG 120
 Query: 121 DKGPVVIKQKLLQKGLSKSIETSELAQINFPDLASKISQKLYKQKQLPLAKLKDLMQ 180
 DKGP V+KQKLLQKGI S I+ L +P LA K+SQKL+ KYQ KLP KALSDK+ Q
 10 Sbjct: 121 DKGFQVIAQKLLQKGLASHDIDPILSQIDFSLAQKVSQKLFUKYQKQLPPKALDKKITQ 180
 Query: 181 SLTFKGPDPYQIVHTVIGQLRIEKDQELERDLVYKHLDWQYQKLSKQHDQVHLQRIINAL 240
 +L TRGF Y + +L ++D + EDL+ KELDKQY+KLS+K+D Y LEQ++ AL
 Sbjct: 181 ALLTKGFSYDLAKHSINHLNFDQDNQETEDLLDKELDRYKLSRAKYDGYTLRQKLYQAL 240
 15 Query: 241 MRKGYQYRDIKSALREYL 258
 RKGY +DI LR YL
 Sbjct: 241 YRKGNSDIDNCKLRYL 258

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 263

A DNA sequence (GBSx0278) was identified in *S. galactiae* <SEQ ID 835> which encodes the amino acid sequence <SEQ ID 836>. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3912 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 30 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 The protein has homology with the following sequences in the GENPEPT database:
 >GP:BA04659 GB:AP001510 unknown conserved protein in B. subtilis
 [Bacillus halodurans]
 35 Identities = 96/175 (54%), Positives = 122/175 (68%)
 Query: 1 MRLPKRGDPITTIQSYKHGSLHRTWRTDMLTKTENALIGVMDHTLVEMDGRWVTREP 60
 M PK G I IQSYKH+GS+HR W +T+VLK T +IG ND LV E+DGR W TREP
 40 Sbjct: 1 MNFPKVGSKIQIQSYKHGSIHRIMEETIVLKGTSKVIVIGNDRIVLKESDGRHNR TREP 60
 Query: 61 AIIVYPHKYWFENILAMIRETVGSVYCNLASPYLLDPEALKYIDYDLDKVPADGEKRLD 120
 AI YF + WEN I MIR G+ +YCNL +P+ D EALKYIDYDLD+KVF D +LLD
 Sbjct: 61 AICYPDSQEWNTIGMIRADGIYFYCNLGTPTFWDEALKYIDYDLDKVPFPMFKLLD 120
 45 Query: 121 VDEYBQKAGQWYPTDIDYILKENVKILVENINENKGPSSSYINIWYKRYLEK 175
 DEY H+ M YP +ID IL+ +V LV WI++ KGP+ ++ WY+R+L+ +
 Sbjct: 121 EDEYAMGRIDMKYPPKIDRILQRSVDLVSNIHQKGPFPAPQFVESWYRFLQVR 175

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 837> which encodes the amino acid sequence <SEQ ID 838>. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 55 bacterial cytoplasm --- Certainty=0.3912 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

-329-

Identities = 155/177 (87%), Positives = 165/177 (92%)

Query: 1 MRLPKSGDFTITQSYKHGDSLERTWRDTMVLKTTENALIGVNDHILVTENDGRRWVTREP 60
 N+LPKSGDFTITQSYKHGDSLERTWRDTMVLKTTENALIGVNDHILVTB+DGRRWVTREP
 5 Sbjct: 1 MRLPKSGDFTITQSYKHGDSLERTWRDTMVLKTTENALIGVNDHILVTESDGRRWVTREP 60

Query: 61 AIVVFHKYWFNIIAMIRFQVSYVSYCNLASPYLDPEALKYIDYDLQVKVFADGKRLLD 120
 AIVVFHKYWFNIIAMIR+ GVSYSYCNLASPY++D EALKYIDYDLQVKVFADGKRLLD
 10 Sbjct: 61 AIVVFHKYWFNIIAMIRNDGVSYSYCNLASPYMDTEALKYIDYDLQVKVFADGKRLLD 120

Query: 121 VDEYEHKKYQNMYPITDIDYILKENVKILVSNIMENKGPSPSSYINIWYKRYLELQGR 177
 VDEY HK +M Y D+D+ILKENVKILV+WIN KGPSS +YI IWYKRYLELK R
 15 Sbjct: 121 VDEYEHKKYQNMYPITDIDYILKENVKILVWINHEKGPSPKAYITIWYKRYLELQGR 177

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 264

A DNA sequence (GBSx0288) was identified in *S. agalactiae* <SEQ ID 839> which encodes the amino acid sequence <SEQ ID 840>. This protein is predicted to be jag protein. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.1666 (Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000 (Not Clear) < succ>
bacterial outside ---	Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB07782 GB:AP001520 spoIIJ-associated protein [Bacillus halodurans]
 Identities = 54/198 (27%), Positives = 98/198 (49%), Gaps = 6/198 (3%)

Query: 100 DVVEEYIEVDTELEKEDVSQPELPKIDDKNVTTSEAIKIDLLPNIEVAAQVTKYVE 159
 + VE+ I E+ T E+ + E PK ++ + A+ ++ + P+ + ++E
 35 Sbjct: 13 EAVEQAIIELGTTRERITYTVVEEPKSLPGILGSKPAVIEVVVKPD---FVDRAKAFLE 69

Query: 160 NIIYEMDLDA--TIETTTSKRQINLIQISTPRAGRIIYHGKVLKSLQLLAQNYLHDPFSK 217
 ++ EMD++ TIE + N+ E + G +IG G+ L SLQ L+ + +
 40 Sbjct: 70 ELLQCMCEVETIEKDPATVLFNLISGQ-DLGTLLGKRGQTLDSLQYLVNLVANKEGSS 128

Query: 218 SFSVSVINVDYVEHRTETLIDFSKKIARVLETNRPYHMDPMSNERKTVHKTIATIEGV 277
 + ++ +Y R E L+ +++A + L T P ++FMS ERK +H + + V
 45 Sbjct: 129 PIRIKLDAENYRARRKEALVOLARLASKALRTKRPVSLFMSAHERKIHTALQELGDV 188

Query: 278 ESYSEGNDPNRFVVVTKK 295
 E+YSEG R VV+ K
 50 Sbjct: 189 ETSYSEGQIGRHVVIAPK 206

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 841> which encodes the amino acid sequence <SEQ ID 842>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.3721 (Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000 (Not Clear) < succ>
bacterial outside ---	Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

-330-

Identities = 176/302 (58%), Positives = 223/302 (73%), Gaps = 32/302 (10%)

Query: 23 MVLPTGATVESAIEKGLRLNISRRAHIKVSREKKGFLGPGKKPAKVEIGTIDVTD 82
 MVLPTG TVESAIE GLRL +SR+AHIKV+S+EKKGFLGPGKKPA+V+IBOI+D+
 5 Sbjct: 1 MVLPTGKTVEAIEKGLRLSLRKAHIKVTSEKKGFLGPGKKPAQVDIEGIDKTVY 60

Query: 83 INESVALKNI-----KNVPS--SVDVREYIEKVDETLEKEDVSPQLPKDDK---- 129
 + A + + +N P+ S DV E I+ + LE ED L D
 10 Sbjct: 61 KADKKAIEGVPEIDINKTPAVNSADVRPERIKAT-QRLEAEDTKVPELMSSEDAQTPS 119

Query: 130 ---NVVITSEA-----IEKIDL-----LPIIEVAAQVTKYVENIIEYMDLADI 171
 VT ++A +E+ ++ +IE AA +V+ YV IIEYMD++AT+
 15 Sbjct: 120 NLAETVIEITKAQQPSIPVEESEVPQDAGNDGFSKDIEKAAQEVSDYVTKIIEYMDIETV 179

Query: 172 ETTTSKROINLQIETPAGRIIGYHGKVLKSLQLLAQNLYLHDFSKSPSVSINVVDYVEH 231
 ET+ ++ROINLQIETPAGR+IGYHGKVLKSLQLLAQN+LHDR+SK+FSVS+NVHDYVEH
 15 Sbjct: 180 ETSNRRROINLQIETPAGRVIGYHGKVLKSLQLLAQNFLHRYSNFVSINVVDYVEH 239

Query: 232 RTEILDFSKKILARRVLEINEPYHMDPMNSERKTVHKTATIEGVBSYSGNDPNRFV 291
 RTEILDFE++K+AAVLE+ + Y MDPMSNSEK VHKT+++IEGV+SYSGNDPNR+V+
 20 Sbjct: 240 RTEILDFETQKVAKRVLESQDQYTMDPMSNSEKIVHKTVSSIEGVDSYSGNDPNRYV 299

Query: 292 VT 293
 V+
 25 Sbjct: 300 VS 301

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 265

- 30 A DNA sequence (GBSx0290) was identified in *S. galactiae* <SEQ ID 843> which encodes the amino acid sequence <SEQ ID 844>. This protein is predicted to be 60 kd inner-membrane protein (yidC). Analysis of this protein sequence reveals the following:

Possible site: 42

>>> May be a lipoprotein

35	INTEGRAL	Likelihood = -7.38	Transmembrane	54 - 70 (52 - 75)
	INTEGRAL	Likelihood = -5.20	Transmembrane	193 - 209 (192 - 211)
	INTEGRAL	Likelihood = -3.61	Transmembrane	125 - 141 (124 - 144)
	INTEGRAL	Likelihood = -2.44	Transmembrane	168 - 184 (167 - 184)

40 ----- Final Results -----

bacterial membrane --- Certainty=0.3951(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 45 The protein has homology with the following sequences in the GENPEPT database:

>GP:CA978595 GB:Z14225 SpoIIIV [Bacillus subtilis]

Identities = 79/243 (32%), Positives = 142/243 (57%), Gaps = 5/243 (2%)

Query: 1 MCKKLTFTSLILLTGLSLVACS--RGEVSSHSMTLWQ-IVYAPAKSIQWLS--FMSIG 55
 MK+++ ++ LL C + +++ S W++ +VY + +1 +++ + G
 50 Sbjct: 1 MRRIRGLLSMVGVMFLAGCSSVKEPTADSPHFWDKYVVYPLSELITVAKLGTGNYG 60

Query: 56 LGIILFTLIRAIMPLYNMCKKSSQRMQRIQPLRKLKQKYPGKDPUNRLKLANDMQSM 115
 L IIL T++IR +++PL Q++SS+ MQ +QP +++t++KY KD +KL E ++
 55 Sbjct: 61 LSIILVTLIRLILLPLNIRKLRSSKAMQALQPEMKLREKYSRQDKYQQLQLQETVAL 120

Query: 116 YKAGVNPYASVLELLIQLFLVIALPQALTRVSPFLKVTFLSLSLSQDPPFLLPLAAL 175
 ++ GNP A E+LIQ+P+L + A+ R + +FL +L+ DPTFLP++A+ +
 60 Sbjct: 121 FQKHGVNPLAGCFPILIQMPLIGFYHAIIMRTQAISHSFLWFDLGEKDDPFFILPVAGV 180

Query: 176 FTFPLSTVLTNKAARVKNITLITMYVMPFILLVTSFNPASGVLYVWVNSAPQVFIILL 235

-331-

TF+ L ++N ++M ++MP +I+V + NF + + LYW V N F + Q L+

Sbjct: 181 ATFVQQLMAGNAGQQNPQMAMLMWIMPIMIIVFAINFPAALSLYVVGRLFMIAQTFLI 240

Query: 236 INP 238
P

Sbjct: 241 KGP 243

A related GBS sequence was identified <SEQ ID 10783> which encodes amino acid sequence <SEQ ID 10784>.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 845> which encodes the amino acid sequence <SEQ ID 846>. Analysis of this protein sequence reveals the following:

Possible site: 49

>>> May be a lipoprotein

INTEGRAL	Likelihood = -6.32	Transmembrane	198 - 214 (197 - 220)
INTEGRAL	Likelihood = -5.52	Transmembrane	59 - 75 (57 - 80)
INTEGRAL	Likelihood = -4.25	Transmembrane	130 - 146 (129 - 150)
INTEGRAL	Likelihood = -2.28	Transmembrane	173 - 189 (170 - 189)

----- Final Results -----

bacterial membrane --- Certainty=0.3527 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BA005234 GB:D26185 stage III sporulation [Bacillus subtilis]
Identities = 90/249 (36%), Positives = 150/249 (60%), Gaps = 6/249 (2%)

Query: 16 IVPVLVLLVACG--RGEVTAQSSSGWDQ-LVYLEFAIAQWLS--FDGSIGVGILITLTI 70

+V ++LL C ++TA S WD+ +VY + I +++ +G+ IIL T+ I

Sbjct: 13 MVGVFMILAGCSSVKEPITADSPHFWDKYVYPLSELITYVAKLTGDNGLSILVITLIL 72

Query: 71 RLMLPLFMIMQIKSSQMODIOPELRELQRKYAGKDTOTRMKLAESQALYKKYGVNPA 130

RL+++PL Q++SS+ MQ +QPE++L+ KY+ KD +T+ KL +E+ AL++K+GVNP A

Sbjct: 73 RLLILPLMIKQLRSSIGAMQALQPEMQLKEKYSKDKTKQKLIQOETMALFQKHGVNPLA 132

Query: 131 SLLEPLLIQMPVMIALFOALTRVSLFKTGTFVLAQHDHLLYLFVLAAVFTLSWLTN 190

P+LIQMP++I +A+ R + +FLW +L+ D Y+LF++A V TF+ L

Sbjct: 133 GCFPIILIQMPILIGFYHAIMRTQAISEHSFLWFLDGEKDPYILFIVAGVATFVQOKLM 192

Query: 191 LAAKEKNVMTVMITYVMPIMIFMGFNLAGGVVLYWTVSNARQVQLLLINP-FKIIAE 249

++N M +M+++MP+MI N ++ LYW V N F + Q L+ P K E

Sbjct: 193 AGNACQNCPQAMLMWIMPIMIIVFAINFPAALSLYVVGRLFMIAQTFLIKGPDIKKNPE 252

Query: 250 RORLANEEK 258

Q+ ++K

Sbjct: 253 PQKAGGKKK 261

An alignment of the GAS and GBS proteins is shown below:

Identities = 172/270 (63%), Positives = 217/270 (79%), Gaps = 1/270 (0%)

Query: 1 MKKKLKTFSILLITGSLLVACURGEVSSHATLWBOIVYPAKSIQWLSFNHISGLGITL 60

+KK +K ++L LLVACURGEV++ S++ W+Q+VY F++I+QWLSF+ SIG+GITL

Sbjct: 7 VKNNIKIARIIVPLV-LLLVACURGEVTAQSSSGWDQVYLFAIAQWLSFDGSIGVGIL 65

Query: 61 FTLIRAIMMFLYNNQMKSSQRMQEIQPRILKELQKKYPGKDPNRLKLNDEMCSMYKAE 120

FTL IR ++MPL+NMQ+KSSQRMQ+IQP L+ELQ+KY GKD R+KL +E Q++YK G

Sbjct: 66 FTLIRILNLMFLNMQ+KSSQRMQEIQPELRELQRKYAGKDTOTRMKLAESQALYKKY 125

Query: 121 VNPIASVLPILLIQPLVLMALFOALTRVSLFKTGTFSLSELSQDPYIILPVLAALFTPLS 180

VNPIAS+LPLLQ+PV+ ALFQALTRVSLFK GTFL +EL+Q D Y+LPLVLA+FTPLS

Sbjct: 126 VNPIASVLPILLIQPLVLMALFOALTRVSLFKTGTFVLAQHDHLLYLFVLAAVFTPLS 185

-332-

Query: 181 TWLTNKAARVEKNIALILNTYVMPFIIIVISPNFASGVVLYTWVSNAPQVQILLINPNFK 240
 TWLTN AA EKH+ +T+M YVMP +I FN ASGVVLYTWVSNAPQV Q+LLINPN+K
 Sbjct: 186 TWLTNLAARKEKRVNMTVMYVMPFIMFPMFGNLAAGVVLYTWVSNAPQVQILLINPNFK 245

Query: 241 IIKVKEEAVRVAHKEKQVRVKRKRKASKKR 270
 II R+ E+ R +RA++KA K++
 Sbjct: 246 IIAERQRLANEKEERKLREARRAKKAMKRK 275

- 10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8527> and protein <SEQ ID 8528> were also identified. Analysis of this protein sequence reveals the following:

15 Lipop: Possible site: 20 Crend: 5
 MG: Discrim Score: 4.90
 GvH: Signal Score (-7.5): -0.35
 Possible site: 42
 >>> May be a lipoprotein

20 ALOM program count: 4 value: -7.38 threshold: 0.0
 INTEGRAL Likelihood = -7.38 Transmembrane 54 - 70 (52 - 75)
 INTEGRAL Likelihood = -5.20 Transmembrane 193 - 209 (192 - 211)
 INTEGRAL Likelihood = -3.61 Transmembrane 125 - 141 (124 - 144)
 INTEGRAL Likelihood = -2.44 Transmembrane 168 - 184 (167 - 184)
 PERIPHERAL Likelihood = 2.54 217

25 modified ALOM score: 1.98

*** Reasoning Step: 3

----- Final Results -----

30 bacterial membrane --- Certainty=0.3951(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

35 32.8/62.3% over 242aa

EGAD|17722| stage III sporulation protein j precursor Bacillus subtilis
 OMNI|NT01BS4782 -identity Insert characterized Insert characterized

40 SP|Q01625|SP3J_BACSU STAGE III SPORULATION PROTEIN J PRECURSOR. Edit characterized
 GP|40023|amb|CAA44401.1||X62539 unnamed protein product Insert characterized
 GP|467388|dbj|BA05234.1||D26185 stage III sporulation Insert characterized
 GP|263665|amb|CA16141.1||Z99124 alternate gene name: spoJ87 Insert characterized
 PIR|I40437|I40437 stage III sporulation protein spoIIJ - Insert characterized

45 ORF02221(301 - 1014 of 1413)
 EGAD|17722| S4098(3 - 245 of 261) stage III sporulation protein j precursor { acillus
 subtilis|OMNI|NT01 84782 -identitySP|Q01625|SP3J ACEU STAGE III SPORULATION PROTEIN J
 PRECURSOR.GP|40023|amb|CAA44401.1||X62539 unnamed protein product { acillus
 subtilis|GP|467388|dbj| AA05234.1||D26185 stage III sporulation { acillus
 subtilis|GP|263665|amb|CA 16141.1||Z99124 alternate gene name: spoJ87 { acillus
 subtilis|PIR|I40437|I40437 stage III sporulation protein spoIIJ - acillus subtilis
 %Match = 17.0
 %Identity = 32.8 %Similarity = 62.2
 Matches = 79 Mismatches = 88 Conservative Sub.s = 71

55 219 249 279 309 339 393 420
 DFVVIARGVVEELDYQALEKNNIHLVLIAGLTI*KGILKKKKLTKFSLILITGSLVACG--RGEVSSHSATLWEQ-IVYA
 :||::: : : : || | : : : | : : : ||
 MLKKRRIGILLSMGVGVFMILAGCSVSKPEITADSPHFVWQVYVP
 10 20 30 40

60 474 504 534 564 594 624 654
 FAKSIQWLS--FNHSIGLGIILFTLIIRATNMPLYNMOMKSSORMOEIQPRILKELQKYQPKDIDNRLKINDQMYSYKA

-333-

[illegible]

Example 266

A DNA sequence (GBSx0291) was identified in *S.agalactiae* <SEQ ID 847> which encodes the amino acid sequence <SEQ ID 848>. Analysis of this protein sequence reveals the following:

```

Possible site: 46
20 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3778(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
25      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9579> which encodes amino acid sequence <SEQ ID 9580> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

30      >GP:CA4440.GB:XE2539 homologous to E.coli rmpA (Bacillus subtilis)
      Identities = 52/109 (47%), Positives = 77/109 (69%), Gaps = 1/109 (0%)

Query: 21  LKTTTYSKSDNDPOMPSRSGKNVANKVPIYLYLEK-EQKHFFVGVISVSKLQVAVNRNAI 79
      LKK  RAK + +DFO + F G +VAVNR+PV+Y L++ E  RVG+SVSKK+GVAV+RN I
35  Sbjct: 4  LKKRIRKKNEDNDPQVFKHG+SVANRGPFVLYTLQPEDELRLVGLSVSKK+GVAVNRNI 63

Query: 80  RKRIRHVLLSKQKTALQDYDFVGVARGVELDQALEKNLHLVGLKIAGL 128
      KR IR L +K L++ D+++IARK +L+Y +K+L++ + +L
      KR LRIQRLFLKKERLKEKYIV+IIRKPSALQVTEETKSL+HLFRKSSL 112
  
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 849> which encodes the amino acid sequence <SEQ ID 850>. Analysis of this protein sequence reveals the following:

```

Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3620 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities 73/109 (66%), Positives = 88/109 (79%)

Query: 21 LKTRVYRKSDIDDPOMIPSRKRVANRKFVYIYLEKQKQVFRVGLSGVKKLGNAVTVRVAIK 80
      LKTRVYRK+KIDQ IF GK+ ANRKFVY+L+Q Q HFRVGLSG KK+GNNAV RMA+K
Sbjct: 1 LKTRVYRKREDDQAIFKQKGGSTANRKFVYIYHLNRQDQHFRVGLSGVKKIGNAVTRVAVK 60

Query: 81 RKRTRVHLKSQTLAQDYPVPIARKGQVEEDLYQALEMKLHLVTKLGI 129
      RKRTRV+++ L+ DPVPIARKGV L+YQ L+L+L+L+L+L+L+L+L+

```

Subjct: 61 RKIRHVIMALGHQLKSEDFVVIARKEGVHSLEVQLQNLHHVLKLAQLL 109

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

5 Example 267

A DNA sequence (GBSx0292) was identified in *S. agalactiae* <SEQ ID 851> which encodes the amino acid sequence <SEQ ID 852>. This protein is predicted to be glycerol-3-phosphate dehydrogenase, NAD-dependent (gpsA). Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1429 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 852> which encodes amino acid sequence <SEQ ID 853> was also identified. There is a signal peptide at residues 1-19. The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA86746 GB:U32164 NAD(P)H-dependent dihydroxyacetone-phosphate
 reductase [Bacillus subtilis]
 Identities = 177/333 (53%), Positives = 241/333 (72%)

Query: 18 QKIAYLPGPSWGTALAQVLMDNGHEVRLAGNVVEICEINTHINQRYPKDITLDSKIKIA 77
 +X+ +LG GSWGTALA VL ENG+EV +W + + I +IN H N+ Y ++ L + IK
 Sbjet: 2 KKVINGAGSGWGTALALVITDNGNEVCVMHAFDLHINELHENVKYLENVGLSTBIIG 61

Query: 78 YTNLEEARINNVDSLLFVVPFTRVLVAQVAMLLKHKVVLVMAHASKGLEPOTHERLSTILE 137
 T+++EA+++ D I+ VPTK R V +Q + K V +H SKG+EP + R+S I+E
 Sbjet: 62 TIDMKERAVSDADVITVAVPTKAIREVLRQAVPPITTKAVPVVHVKGIKIEPDSLLRSEIME 121

Query: 138 EEISEQYRSDIVVVSGPSHAEATVRDITLITAASKDIEAAKVKVQLFNSHYFRLYNTND 197
 E+ R DIVV+SGPSHAE +R T +TA+SK + AA+ VQ LF NH FR+YTN D
 Sbjet: 122 IELPSDVRDITVVLGSPSHAEVQLRHATTYTTASSKEMRAAEVQDLFINHFRVYNTND 181

Query: 198 VGVGTAGRLNIIIAVGGALHGLGYCDNNAKAAIITRGLAEITRGLGVGLGADPLTPSGLS 257
 ++GVE QRLNIIIA+ AG GLGYCDNNAKAAIITRGLAEI RLG ++G +PLTPSGL+
 Sbjet: 182 IIGVEIGRLNIIIAIAGITDGLGYCDNNAKAAIITRGLAEIARLGTGMGNPLTPSGLT 241

Query: 258 GVGDILVTCTSVHSRNWRAGDALGCGEKLEDIEKRMCMVIEGISTTKVAYEIAQNLNVYM 317
 GVGDILVT TSVHSRNWRAG+ LG+G KLERD+ + MCMV+EG+ TTK AY++++ +V M
 Sbjet: 242 GVGDILVTCTSVHSRNWRAGNLGKGYKLEDVLEEMCMVVEGVRTTKAAYQLSKKIDVKM 301

Query: 318 PITEAIYKSIYEGANIKDSILDWMSNEPSENR 50
 PITEA+++ ++ G ++ ++ +M+ E E
 Sbjet: 302 PITEALHQLVFGQKQVETAVESLMAGKTHMB 334

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 853> which encodes the amino acid sequence <SEQ ID 854>. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0882 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

-335-

Identities = 287/338 (84%), Positives = 316/338 (92%)

Query: 15 MTKQKIAVLGPQSWGTALAQVLNDNGHSEVRLGNIVVDQIEEININHTNQRYPKIDITLDSK 74
 5 Sbjct: 1 MTKQKIAVLGPQSWGTALAQVLNDNGHSEVRLGNIVVDQIEEININHTNQRYPKIDITLDSK 60

Query: 75 IKAYINLSEAINIVDSILPVVPTKVTRIAKQVANLLKHVVLMIAASKGLEPCTHERLST 134
 I A +L +A+++VD++LPVVPVKVTRIA+QVA +L HKVV+MIASKGLEPCTHERLST
 10 Sbjct: 61 ITATLDLQALSDVDVLPVVPVKVTRIAQVAAILDHKVV+MIASKGLEPCTHERLST 120

Query: 135 ILEEEISQYRSDIVVSGPSHAEAIIVRDITLITASKDIEAAKYVQKLSNHYFRLYT 194
 ILEEEI +RS++VVVSGPSHAE IVRDITLITASKDIEAAKYVQ LPSNHYFRLYT
 10 Sbjct: 121 ILEEEIPAFRSEVVVVSGPSHAEITVRDITLITASKDIEAAKYVQSLPSNHYFRLYT 180

Query: 195 NTDVVGVEGTAGALNNTIYVAGALHGLGVGDNAKAAIITRGLAEITRIGVOLGADPLTFS 254
 NTDV+GVETAGALNNTIYVAGALHGLGVGDNAKAA+ITRGLAEITRIGVOLGADPLT+FS
 15 Sbjct: 181 NTDVVGVEGTAGALNNTIYVAGALHGLGVGDNAKAAVITRGLAEITRIGVOLGADPLTFS 240

Query: 255 GLSGVGLDIVGTSTVHSRNRWAGDALGRCKLEDIERNMGMVIEGISTTKVAYEIAQNLN 314
 GLSGVGLDIVGTSTVHSRNRWAG ALGRCKLEDIE+NMGMVIEGI+TTKVAYEIAQ+L
 20 Sbjct: 241 GLSGVGLDIVGTSTVHSRNRWAGALGRCKLEDIERNMGMVIEGIATTKVAYEIAQDLG 300

Query: 315 VYMPITTAIYKSIYEGANIKSDILDMMSNFRSENEWH 352
 VYMPIT IYKSIYBGA+IK+SL MMSNFRSENEWH
 25 Sbjct: 301 VYMPITTAIYKSIYEGADIKESILDMMSNFRSENEWH 338

SEQ ID 8530 (GBS291) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 59 (lane 5; MW 38.9kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 2; MW 64kDa).

30 GBS291-GST was purified as shown in Figure 226, lane 10-11.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 268

A DNA sequence (GBSx0293) was identified in *S. agalactiae* <SEQ ID 855> which encodes the amino acid sequence <SEQ ID 856>. This protein is predicted to be glucose-1-phosphate uridylyltransferase (gtaB). Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have a cleavable N-term signal seq.

40 ----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database:

>GP:BAA28714 GB:AB001562 glucose-1-phosphate uridylyltransferase
 [Streptococcus mutans]
 Identities = 263/296 (88%), Positives = 285/296 (95%)

50 Query: 2 KVRKAVIPANGLGSTRFLPATKALAKHMLPIVDKPTIQFIVEEALKSIGIEDILVVTGSKR 61
 KVRKAVIPANGLGSTRFLPATKALAKHMLPIVDKPTIQFIVEEALKSIGIEDILVVTGSKR
 Sbjct: 5 KVRKAVIPANGLGSTRFLPATKALAKHMLPIVDKPTIQFIVEEALKSIGIEDILVVTGSKR 64

55 Query: 62 SIEDHFDNFELEYNLKKGKNNLLKLVDETTGIRLHPIRQSHFRGLGDVILQAKPVGN 121
 SIEDHFDNFELEYNL++KKG +LLKLV++TT I LHFIRQSHFRGLGDVILQAKPVGN
 Sbjct: 65 SIEDHFDNFELEYNLKKGKNTDLLKLVNDITAINLHPIRQSHFRGLGDVILQAKPVGN 124

Query: 122 EFPVVMGLGDLDYITNKVIPLTKOLINDFRATHASTIAVMVEPKEDVSAYGVIAFQGB 181

-336-

```

EPFVVMIGDDLMDIT++K IPLT+QL+ND+E THASTIAVMKVPHSDVSAYGVIA PQGG
Sbjct: 125 EPFVVMIGDDLMDITDDICAIPLTQRLMDYESTHASTIAVMEVPHSDVSAYGVIA PQGG 184

Query: 182 VNGLYSVNTFVEKPSPEAPSNLAIIIGRYLLTPEIPNILETKQPGAGNEIQLTDAIDTLN 241
V+GLYSV+TFVEKP+P+EAAPSNLAIIIGRYLLTPEIF ILESTQ+PGAGNE+QLTDAIDTLN
Sbjct: 185 VSGLYSVDTFVEKPAKPSAPSNLAIIIGRYLLTPEIFILESTQPGAGNEVQLTDAIDTLN 244

Query: 242 KTORVFARKFTGDRYVDGDKFGPMKTSIDYALHPQVKKDLKYYIIDLGSLEKTS 297
KTORVFAR+P G RYVDGDKFGPMKTSIDYAL+HPQVK+DLK YII+LGR L++ S
Sbjct: 245 KTORVFAREFKGRYVDGDKFGPMKTSIDYALGHPQVKEDLKAYIIDLGRKLQKS 300

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 857> which encodes the amino acid sequence <SEQ ID 858>. Analysis of this protein sequence reveals the following:

```

Possible site: 26
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 257/295 (87%), Positives = 277/295 (93%)

Query: 2 KVRKAVIPAAGLGTRFLFATKALAKEMLPIDVKPTIQFIVEALKSGIEDILNVTKSKR 61
KVRKA+IPAAGLGTRFLFATKALAKEMLPIDVKPTIQFIVEALKSGIE+ILNVTK+KR
Sbjct: 3 KVRKAIIPAAGLGTRFLFATKALAKEMLPIDVKPTIQFIVEALKSGIEIILNVTKAKR 62

Query: 62 SIEDHFDNSFELEYNLQKGNELLKLVDETTGIRLHFIRQSHPRGLGDVLAQAKAFVN 121
SIEDHFDNSFELEYNL+ KGNELLKLVDETT I LHFIRQSHPRGLGDVLAQAKAFVN
Sbjct: 63 SIEDHFDNSFELEYNLQKGNELLKLVDETTAIIHFIRQSHPRGLGDVLAQAKAFVN 122

Query: 122 EPFVVMIGDDLMDITBNKVIPLTQQLINDPEATHASTIAVMKVPHSDVSAYGVIA PQGG 181
EPFVVMIGDDLMDITN PLTQQL+ D++ THASTIAVM+VPHSDVS+YGVIA PQG+
Sbjct: 123 EPFVVMIGDDLMDITNASAKPLTQQLMEDYDKTHASTIAVMKVPHSDVSAYGVIA PQGKA 182

Query: 182 VNGLYSVNTFVEKPSPEAPSNLAIIIGRYLLTPEIPNILETKQPGAGNEIQLTDAIDTLN 241
V GLYSV+TFVEKP PE+APS+LAIIIGRYLLTPEIF ILE Q PGAGNE+QLTDAIDTLN
Sbjct: 183 VKGLYSVDTFVEKPEAPSDLAIIIGRYLLTPEIPGILERQTPGAGNEVQLTDAIDTLN 242

Query: 242 KTORVFARKFTGDRYVDGDKFGPMKTSIDYALHPQVKKDLKYYIIDLGSLEKTS 296
KTORVFAR+P G+RYVDGDKFGPMKTSIDYAL+HPQVK+DLK YII LGR+LEK+
Sbjct: 243 KTORVFAREFKGRYVDGDKFGPMKTSIDYALGHPQVKEDLKYYIIDLGRKLEKS 297

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 269

A DNA sequence (GBSx0294) was identified in *S. galactiae* <SEQ ID 859> which encodes the amino acid sequence <SEQ ID 860>. Analysis of this protein sequence reveals the following:

```

Possible site: 42
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -4.94 Transmembrane 28 - 44 ( 27 - 45)

----- Final Results -----
bacterial membrane --- Certainty=0.2975 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

-337-

>GP:CB15143 GB:Z99120 similar to ABC transporter (lipoprotein)
[Bacillus subtilis]
Identities = 148/346 (42%), Positives = 222/346 (63%), Gaps = 16/346 (4%)

5 Query: 31 LITLLSLVSLTLTACGNSDEKAN---KSDIKVAMVTNQGVDDKSFNQSAWBGKQKWK 87
+L+ + L ACGN S + K+ VAMVT+ GGVDDKSFNQSAWBG+Q +GK+
Sbjct: 1 MSLVIAAGTILGACGNSSEKSGSAGBKPKPSVAMVTDVGGVDDKSFNQSAWBGILQAPGKE 60

10 Query: 88 KGLTKG-NGPDFYQSNBSEHDANNLDTAASGGYNLI FGIGFGLHDTIEKVSENNKKVKYV 146
GL KG NG+DY QS +++D+ NL+ A ++LI+G+G+ +D+I +++++ K+ +
Sbjct: 61 NGLKKGKNGYDLQSKSDLDYTNLNLKLAENFDLITGVGYLMEDSISELADQRKNTNFA 120

Query: 147 IVDDIINGKENVASVTFADNERAYLAGVAAAKITTKTKVGFPGCNBGGVVKRPFAGFKMG 206
I+D ++ K+NVAS+TF + E ++L GVAAR ++K+ +GF+GNE ++K+FE GF+AG
15 Sbjct: 121 IIDAIVD-KDNVASITFKQSGSFLVGVAAALBSKSGKIGFVGNBSRLIKKFEVGFAG 179

Query: 207 VESIDPAIKVAVSYAGSFTDAKKGKTIATQYATGVVDVIYQAACOTGNGIFSEAKTENET 266
V++++P V V YAG F A GK A + Y +GVNVIY +AG TG G+P+SAK
Sbjct: 180 VQAVNPKAVVEVKTAGGFDKADVGKATASEMYKSGVDVIYHSAGTITGVVTEAK--NL 236

20 Query: 267 KRESNK--VWVIGVDRDQSGQBNYSKDGKAMFLASTIKEVSGELQSVAEITKQKQP 324
+KE K VRVIGVD+DQ EG +G N L E +K+V ++ V + ++P
Sbjct: 237 KEGDPKRDVWVIGVDRDQYABQV---EGTDNVTITSMVKKVDVTYVEDVTKKASDGKFP 293

25 Query: 325 GCKVTVVGLKDGVDI--KEMLSEBSGVAVKAKEDIVSGKIQVP 368
GG+ +GL GV I + L8 + AV K K+ I+ G +++P
Sbjct: 294 GGETLTYGLDQDQGVGISPSKQNLDDVDIKAVDKKKIIDG-LEIP 338

There is also homology to SEQ ID 862.

- 30 A related GBS gene <SEQ ID 8531> and protein <SEQ ID 8532> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: 20 Crend: 3
Sequence Pattern: CGNR
SRCFLG: 0
35 McG: Length of UR: 19
Peak Value of UR: 2.31
Net Charge of CR: 2
McG: Discrim Score: 5.09
Gvfi: Signal Score (-7.5): -3.29
40 Possible site: 19
>>> May be a lipoprotein
Amino Acid Composition: calculated from 21
ALOM program count: 0 value: 5.20 threshold: 0.0
PERIPHERAL Likelihood = 5.20 90
45 modified ALOM score: -1.54

*** Reasoning Step: 3

----- Final Results -----

- 50 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

- 55 52.8/73.9% over 239aa
Listeria
monocytogenes
SP[Q48754] CD4+ T CELL-STIMULATING ANTIGEN PRECURSOR. Insert characterised
GP[7240601]gb|AAB35725.2||S80336 CD4+ T cell-stimulating antigen Insert characterised
60 CRP02225(385 - 1086 of 1710)
SP[Q48754] TCBSA_L18MO(8 - 247 of 268) CD4+ T CELL-STIMULATING ANTIGEN
PRECURSOR.GP[7240601]gb|AAB35725.2||S80336 CD4+ T cell-stimulating antigen [Listeria
monocytogenes]

-338-

```
%Match = 21.7
%Identity = 52.7 %Similarity = 73.8
Matches = 125 Mismatches = 59 Conservative Sub.s = 50
```

[illegible]

A related GBS nucleic acid sequence <SEQ ID 10947> which encodes amino acid sequence <SEQ ID 10948> was also identified.

SEQ ID 8532 (GBS108) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 7; MW 39.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 9; MW 64.6kDa).

The GBS108-GST fusion product was purified (Figure 202, lane 9) and used to immunise mice. The 35 resulting antiserum was used for FACS (Figure 273), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 270

40 A DNA sequence (GBSx0295) was identified in *S.agalactiae* <SEQ ID 863> which encodes the amino acid sequence <SEQ ID 864>. Analysis of this protein sequence reveals the following:

```

Possible site: 35
>>> Seems to have a cleavable N-term signal seq.
INTEGRAL Likelihood = -12.74 Transmembrane 206 - 222 ( 197 - 224)
INTEGRAL Likelihood = -3.72 Transmembrane 174 - 190 ( 171 - 194)
INTEGRAL Likelihood = -3.19 Transmembrane 98 - 114 ( 96 - 116)
INTEGRAL Likelihood = -1.54 Transmembrane 120 - 136 ( 120 - 139)
INTEGRAL Likelihood = -0.90 Transmembrane 157 - 173 ( 157 - 173)

----- Final Results -----
bacterial membrane --- Certainty=0.6095 (Affirmative) < success
bacterial outside --- Certainty=0.0000 (Not Clear) < success
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < success

```

55 The protein has homology with the following sequences in the GENPEPT database:

-339-

>GP:CA90755 GB:AJ400707 hypothetical protein [Streptococcus uberis]
 Identities = 126/218 (57%), Positives = 166/218 (75%)

5 Query: 8 KEYPTTVLLVSLTPTLAVFLMQIYQSQARSSQVIFQPGGIGDYLKAYPTNMLRISPIF 67
 KR P T +S+T L+F+MQ+ YGS A+S QV+PQPG+ G +K+ P+ LMRLL+PIF
 Sbjct: 5 KEKPVTFPFLSVTILLFIVMQVFFYGSMAKSPQVVFQPGMPLAVKSMPSQLMRLLATPIF 64

10 Query: 68 VHIQWEHPLNGLALYFVGQMGESIWGSLRFLILYLISGLMGNITFLFPTPHVVAAGAST 127
 +HIQWEHPL+N L LYFVGQ+ ESIWGS FL+LY+LSG+MGN+ TLFPPTPHVVAAGAST
 Sbjct: 65 IHIGWEHPLINSITLYFVGQLABSIWGRFPFLLYVLSCIMGNVLTFLFPTPHVVAAGAST 124

15 Query: 128 SLPGVPSAIALMGYPGKMYLQVKGESYQVMILLNLFNIFTPGVSLAGHVGGLVGGVLV 187
 SLPG+P+AI + GYPG N LK +GKESYQ +I+NL N+F P V + GH+GG +GG L
 Sbjct: 125 SLPGLPAAITVVQYFGHQLLESIGESYQTLIILNLVNNLFPMPVGIHVGLGALGGALA 184

Query: 168 AIFLTKQNGSLFLKTIQSLAIAMIFIIVISLIGLSLV 225
 A+FL + LF Q AL+ ++ ++ L LSL+
 Sbjct: 185 AVFLPTLLDAELFTKKQKTSALLSYLTALVLTLSLM 222

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 865> which encodes the amino acid sequence <SEQ ID 866>. Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence

25 INTEGRAL Likelihood = -9.92 Transmembrane 214 - 230 (212 - 232)
 INTEGRAL Likelihood = -5.36 Transmembrane 135 - 151 (128 - 153)
 INTEGRAL Likelihood = -1.81 Transmembrane 101 - 117 (100 - 117)
 INTEGRAL Likelihood = -1.44 Transmembrane 183 - 199 (182 - 199)
 INTEGRAL Likelihood = -0.53 Transmembrane 166 - 182 (166 - 182)

30 ----- Final Results -----
 bacterial membrane --- Certainty=0.4970 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

35 The protein has homology with the following sequences in the databases:

>GP:CA90755 GB:AJ400707 hypothetical protein [Streptococcus uberis]
 Identities = 72/128 (56%), Positives = 94/128 (73%)

40 Query: 106 FLLLYVLSCVMGNAFTFPLTPETVAAGASTSLPGLFAAIVVLSFLGKIQALKDLSKSYQT 165
 FLLLYVLSC+MGN T + TP VAAGASTSLPGLFAAIVV+ + G NQ LK +GKSYQT
 Sbjct: 95 FLLLYVLSCIMGNVLTFLFPTPHVVAAGASTSLPGLFAAIVVVGFGHQLKLSIGKSYQT 154

Query: 166 LIVVNLNLFMPNVSMAHIGGVGGALLSIVFPTKRVITVVKTKRMIALVSYGIIIV 225
 LI+NL+NLNLFMPNV + GH+GG +GGAL+ PT + K ++ AL+SY ++
 45 Sbjct: 155 LIILNVLNLFMPNVGIVGHGALGGALA+VFLPTLLDAELFTKKQKTSALLSYLTAL 214

Query: 226 GVLVIGFL 233
 ++ L +
 Sbjct: 215 VLITLSLM 222

50

An alignment of the GAS and GBS proteins is shown below:

Identities = 63/132 (47%), Positives = 92/132 (68%)

55 Query: 94 GSLRFLLYLISGLMGNITFLFPTPHVVAAGASTSLPGVPSAIALMGYFGKNFYLRQVKG 153
 G FL+LY+LY+MGN FT + TP VAAGASTSLPG+P+AI + + GKN LK +GK
 Sbjct: 102 GLTPFLLYLISGLVMGNAFTFPLTPETVAAGASTSLPGLFAAIVVLSFLGKIQALKDLSK 161

Query: 154 SYQVMILLNLFNIFTPGVSLAGHVGGLVGGVLAIVLPI/TKQNGSLFLKTIQSLAIAMIFI 213
 SYQ +I+NL N+F P VS+GH+GG+VGG L+I + + K + +LAL+ +
 60 Sbjct: 162 SYQTLIVNNLNNLFPMPNVSMAHIGGVGGALLSIVFPTKRVITVVKTKRMIALVSYG 221

Query: 214 IVSISLIGLSLV 225
 I+ + ++ L +
 Sbjct: 222 IILVGVVIGFL 233

A further corresponding DNA sequence was identified in *S.pyogenes* <SEQ ID 9083> which encodes the amino acid sequence <SEQ ID 9084>. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -7.70 Transmembrane 12 - 28 (7 - 30)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4079(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS sequences follows:

Score = 74.5 bits (180), Expect = 5e-16
 Identities = 37/96 (38%), Positives = 48/96 (49%)
 Query: 1 MTQLLRKPYKXXXXXXXXXXXXXAMQVYVGHATGAQAITYGVGGMPGLLVKAMPDQWLRL 60
 M + K YP MQ+ YG A +Q I+Q GG+ G +KA P LKRL
 20 Sbjct: 3 MKKFAKGYPTTVLLVSLVTLVFLMLQLTYGSAESSQVYQFGGIGQDYLKAYPTNLKRL 62
 Query: 61 VTPTXXXXXXXXXXVKNLTLTYFVQGIQVEDLWGSRLF 96
 ++P +NGL LYFVQG+ E +WGS F
 25 Sbjct: 63 ISPIFVHIGWHEFLVGLALTYFVQGMGSIWGSRLF 98

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 271

A DNA sequence (GBSx0296) was identified in *S.agalactiae* <SEQ ID 867> which encodes the amino acid sequence <SEQ ID 868>. Analysis of this protein sequence reveals the following:

Possible site: 22
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2055(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BA028715 GB:AB001562 hypothetical protein [Streptococcus mutans]
 Identities = 96/173 (55%), Positives = 129/173 (74%)
 Query: 1 MEKKILAKKIVLITLKSQPAQYKSEVDCKLLAFIKTKAYQNSCVIATYLSFDYKNTQLL 60
 M KK R +V+ LK Q +A K D +LLE I+ +AYQ + VIATYL+V +E+T L L
 45 Sbjct: 1 MMKKDYRTQIVIRLKKQDKAKKVLIRDRQLLEELIQLEAYQKARVIATYLAPEFSPDLSLL 60
 Query: 61 IKQALDNGKRVLVPKTYPKGMIFVDYQKINIKATITFPGLLEPVNDRVAHVKASIDLHVPG 120
 I+QA D K +VPKTY+ KMIFV Y + +L+ T FGL EP + +A+EK++IDLHVPG
 50 Sbjct: 61 IEQAQINKNSIVVPKTYPRKMIFVVDYADAIQITKPKGLKRPSEKLEKSAIDLHVPG 120
 Query: 121 LIPNNGKFRIGYGAGYFDRLYSDFEGDTISTIVRCQKQDFVEKHDVAIVEVL 173
 L FNN+G+RIG+GAGY+D+YL+DF+GDT+STTY Q+ F D+ VKEVL
 Sbjct: 121 LAFTNNGYRIGPGAGYDQYLDQFQGVVSTIYSPQQIFEPSPFDIPVKEVL 173

A related GBS nucleic acid sequence <SEQ ID 10925> which encodes amino acid sequence <SEQ ID 10926> was also identified.

-341-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 272

- 5 A DNA sequence (GBSx0297) was identified in *S.galactiae* <SEQ ID 869> which encodes the amino acid sequence <SEQ ID 870>. Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence
 10 INTEGRAL Likelihood = -1.44 Transmembrane 161 - 177 (161 - 177)
 INTEGRAL Likelihood = -0.22 Transmembrane 29 - 45 (28 - 45)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1574 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 15 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9305> which encodes amino acid sequence <SEQ ID 9306> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

20 >GP:AAD33517 GB:AF132127 glucose-6-phosphate isomerase
 [Streptococcus mutans]
 Identities = 344/401 (86%), Positives = 374/401 (93%)
 25 Query: 1 MDLPENYDKKEFSRIQKAERKIKSDSEVLVVGIGG3SYLGAKAADIFLNNEFANLQTAEE 60
 ++LP+NYDKKEEF+RI+KAAEKIKSDSEVLVVGIGG3SYLGA+AAIDFIN+ F NL+ EE
 Sbjct: 49 LNLQFNQYDKKEFARIKAAEKIKSDSEVLVVGIGG3SYLGAARAIDFLNNEFANLEKNE 108
 Query: 61 RKAPQILYAGNISISSTYLADLVEYVDQKPSFVNVISKSGTTTEFAIAFRVFKLLVKKYCG 120
 RKAPQILYAGNISISSTYLADLVEYVDQKPSFVNVISKSGTTTEFAIAFRVFKLLVKKYCG
 30 Sbjct: 109 RKAPQILYAGNISISSTYLADLVEYVDQKPSFVNVISKSGTTTEFAIAFRVFKLLVKKYCG 168
 Query: 121 QEEANRIYATTDKVGAVKVEADANWMTFVVPDVGGRFSLTAVGLLPIAASGADIT 180
 QEEANRIYATTDKVGAVKVEADANWMTFVVPDVGGRFSLTAVGLLPIAASGADIT
 35 Sbjct: 169 QEEANRIYATTDKVGAVKVEADANWMTFVVPDVGGRFSLTAVGLLPIAASGADIT 228
 Query: 181 ALMEGAMAARKDLSSDKISENIAYQAAVENVLRKGYITEILANYPESLQFFGEWKKQL 240
 LM GA AAR+D SS ++SEN AYQAA+RN+LYRKY+TE+LANYPESLQFF GEWKKQL
 Sbjct: 229 QLMAGEAARQDYSSALSENENAYQAAIDNLLYRKYITEILANYPESLQFFGEWKKQL 288
 40 Query: 241 AGESBGKDKQGIYPTSANFSTDLHSLQGFICGNNILFETVVRVKPKPKNVITIKLTKDL 300
 AGESBGKDKQGIYPTSANFSTDLHSLQGFICGNNILFETVVRVKPKPKNVITIKLTKDL
 Sbjct: 289 AGESBGKDKQGIYPTSANFSTDLHSLQGFICGNNILFETVVRVKPKPKNVITIKLTKDL 348
 Query: 301 DGLYLQKGDVDFVNRKATDGVLLAHTDGGVPMFVILTPQDAYTIGTYIYFFELAIGLS 360
 DGL YLQKGDVDFVNRKATDGVLLAHTDGGVPM F+T+P QD +TIGY IYFFELAIGLS
 45 Sbjct: 349 DGLAYLQKGDVDFVNRKATDGVLLAHTDGGVPMFVILTPQDAYTIGTYIYFFELAIGLS 408
 Query: 361 GYLNSVNPFDQGVZATKRNMFALLGKFGFEELGAEINARL 401
 GYLN VNPFDQGVZATKRNMFALLGKFGFEEL GAEINARL
 50 Sbjct: 409 GYLNVNPFDQGVZATKRNMFALLGKFGFEELGAEINARL 449

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 871> which encodes the amino acid sequence <SEQ ID 872>. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have no N-terminal signal sequence
 55 INTEGRAL Likelihood = -1.44 Transmembrane 209 - 225 (209 - 225)

-342-

INTEGRAL Likelihood = -0.22 Transmembrane 77 - 93 (76 - 93)

----- Final Results -----

5 bacterial membrane --- Certainty=0.1574 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

10 >GP:AAD33517 GB:AF132127 glucose-6-phosphate isomerase
[Streptococcus mutans]
Identities = 369/449 (82%), Positives = 408/449 (90%)

Query: 1 MSHITFDYKSVLESFAGQREIDFLQGVTRADKLLRBOTGPGSDFLQWLDPENYDKDEF 60
M+HI FDYKSVL F HE+D++Q QVT AD+ LR+GVGPG++ GML+LP+NYDK+EF
15 Sbjct: 1 MTHIKFDYKSVLKGKFLASHELDTYQMQVTAADENLRKGTGGAENTGWLNLPCNYDKKEF 60

Query: 61 ARILTAARKIKADSEVLVVIIGGSYLGAKAIDFLNNHFANLQTAKEKRAQPILYAGNS 120
ARI AAEKIK+DSEVLVVIIGGSYLGA+RAIDPLN F NI+ +ERKAPQIILYAGNS
20 Sbjct: 61 ARIKKAARKIKSDSEVLVVIIGGSYLGAARAIDPLNSSFVNLENKEKRAQPILYAGNS 120

Query: 121 ISSITLADLVEYVQDKEPSVNIVSKSGITTEPAIAFRVFKLLLVKKYQGBEANKRIYATT 180
ISS YLADL+YV DK+PSVNIVSKSGITTEPAIAFRVFK+LLVKKYQGBEANKRIYATT
25 Sbjct: 121 ISSNYLADLVDYVADIK+PSVNIVSKSGITTEPAIAFRVFKDLLLVKKYQGBEANKRIYATT 180

Query: 181 DKVGAVKVEADANNWETFFVVDNVGGRFSVLTAAGLLFAASGADITALMSGANNAARD 240
D+VGAVKVEADAN WETFFVVD+VGGRF+VLTAAGLLFAASGAD+ LM GA AAR+D
30 Sbjct: 181 DRVKGAVKVEADANGWETFFVVDNVGGRFTVLTAAGLLFAASGADLDQLNAGAEAAARD 240

Query: 241 LSSDKISENIAQYAAENVNLYRKYITEILANYEPSLQYFOEWMWQLAGESEBKDQKGI 300
SS ++SEN AYQYAA+RN+LYRKY+TE+LANYEPSLQYF EWMWQLAGESEBKDQKGI
35 Sbjct: 241 YSSAEISENIAQYAAENILNLYRKYITEVLANYEPSLQYFSEWMWQLAGESEBKDQKGI 300

Query: 301 YPTSANFSTDLHSLGQFIQBGSRNLFETVIRVDNFRNVIIPELASDLGLGLYLGQKDV 360
YPTSANFSTDLHSLGQFIQBG RNLFTVIRV+ RIN++++PE ASDLDGL YLGQKDV
40 Sbjct: 301 YPTSANFSTDLHSLGQFIQBGSRNLFETVIRVEKARNILVPEASDLGLGLYLGQKDV 360

Query: 361 FVNKKATDGVLLAHTDGGVFNFMVTLPAQDEFTLGTYTFYFELAIASGVN+GAVNPFDPQ 420
FVNKKATDGVLLAHTDGGVFN F+T+P QDEFTLGY IYFELAI +SGV+N VNPDPQ
45 Sbjct: 361 FVNKKATDGVLLAHTDGGVFNFTLIT+PEDEFTLGTVIYFELAIQLSGVYLGAVNPFDPQ 420

Query: 421 GVEAYKKNFALLGKPGFEALSABLWRL 449
GVEAYK+KNFALLGKPGFE L ABLWRL
50 Sbjct: 421 GVEAYKKNFALLGKPGFEALGABLWRL 449

45 The protein has homology with the following sequences in the databases:

>GP:CA890755 GB:AJ400707 hypothetical protein [Streptococcus
uberis]
Identities = 58/91 (63%), Positives = 69/91 (75%)

50 Query: 6 KRYPTITFLLGLTGLIFIAMQVYVGHATGAQAIYQVGGMFGLLVKMPDQLRWLVTPIF 65
K P+T F L +T L+PI MQV YG A Q++Q GGMFGL+VK+MP QLWRLVTPIF
Sbjct: 5 KEKPTVTFVFLSVTILLPIVMQVYFGSNARSPQVFPQGMFGLVVKMSPQLRWLVTPIF 64

Query: 66 IHIGFGRHFPVNGLTLYFVGQIVEDLWGRSLF 96
IHIG+ HF +N LTLFYVGQ+ E +WGR F
55 Sbjct: 65 IHIGWEHFLNLSLTLYFVGQLAESWGRSFF 95

An alignment of the GAS and GBS proteins is shown below:

Identities = 380/401 (94%), Positives = 392/401 (96%)

60 Query: 1 MDLPENYDKKEPSRIQKAAEKIKSDSEVLVVIIGGSYLGAKAIDFLNNHFANLQTAEE 60
+DLPENYDK+EF+RI AAEKIK+DSEVLVVIIGGSYLGAKAIDFLN+HFANLQTA+E
Sbjct: 49 LDLPENYDKDEPARILTAARKIKADSEVLVVIIGGSYLGAKAIDFLNNHFANLQTAGK 108

-343-

Query: 61 RKAPQLIYAGNISISSTYLADLVEYVQCKEFSVNVISKSGTTTEPAIAFRVFKELLVKVKG 120
 RKAPQLIYAGNISISSTYLADLVEYVQCKEFSVNVISKSGTTTEPAIAFRVFKELLVKVKG
 Sbjct: 109 RKAPQLIYAGNISISSTYLADLVEYVQCKEFSVNVISKSGTTTEPAIAFRVFKELLVKVKG 168

5 Query: 121 QREANKRIYATTDKKVEGAVKVEADANWETVVPVNVGGRFSVLTAAGLLPIAASGADIT 180
 QREANKRIYATTDKKVEGAVKVEADANWETVVPVNVGGRFSVLTAAGLLPIAASGADIT
 Sbjct: 169 QREANKRIYATTDKKVEGAVKVEADANWETVVPVNVGGRFSVLTAAGLLPIAASGADIT 228

10 Query: 181 ALMEGANAARKDLSSDKISENIAQYAAVRNVLYRKGYITTEILANYEPSLQYFGEWKQL 240
 ALMEGANAARKDLSSDKISENIAQYAAVRNVLYRKGYITTEILANYEPSLQYFGEWKQL
 Sbjct: 229 ALMEGANAARKDLSSDKISENIAQYAAVRNVLYRKGYITTEILANYEPSLQYFGEWKQL 288

15 Query: 241 AGESEGGDKQGIYPTSANPSTDLSLGGQFIQSGYRNLFTVVRVEKPRKNVITPELFDL 300
 AGESEGGDKQGIYPTSANPSTDLSLGGQFIQSGYRNLFTVVRVEKPRKNVITPELFDL
 Sbjct: 289 AGESEGGDKQGIYPTSANPSTDLSLGGQFIQSGYRNLFTVVRVEKPRKNVITPELFDL 348

20 Query: 301 DGLGYLGGKDVDFVNGKATDGVLLARTDGGVPMFVTLPTQDANTLYGTYIYFFELAIGLS 360
 DGLGYLGGKDVDFVNGKATDGVLLARTDGGVPMFVTLPTQDANTLYGTYIYFFELAIGLS
 Sbjct: 349 DGLGYLGGKDVDFVNGKATDGVLLARTDGGVPMFVTLPTQDANTLYGTYIYFFELAIGLS 408

Query: 361 GYLSNVNPFDDQPGVEAYKRNMFALLGKPGFELSSELNARL 401
 GY+N+VNPFDQPGVEAYKRNMFALLGKPGFELSSELNARL
 Sbjct: 409 GYMNANVNPFDQPGVEAYKRNMFALLGKPGFELSSELNARL 449

- 25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 273

A DNA sequence (GBSx0298) was identified in *S.galactiae* <SEQ ID 873> which encodes the amino acid sequence <SEQ ID 874>. Analysis of this protein sequence reveals the following:

30 Possible site: 38
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.66 Transmembrane 654 - 670 (653 - 671)
 INTEGRAL Likelihood = -1.65 Transmembrane 113 - 129 (113 - 129)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.2062 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 40 A related GBS nucleic acid sequence <SEQ ID 9463> which encodes amino acid sequence <SEQ ID 9464> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA81906 GB:D04863 alcohol dehydrogenase 2 [Entamoeba
 histolytica]
 45 Identities = 536/864 (62%), Positives = 663/864 (76%), Gaps = 3/864 (0%)

Query: 20 ETTDVALAIDTLVQKGLKALDEMR--QLNQEQVDYIVAKASVAALDAHGLALHAVERGT 77
 +T V I+ LV+ AL E + QE+DYIV KASVAALD H LA AVSEGT
 Sbjct: 5 QTMVDEHINQLVRKAQVALKEYLPEYTOEKIDYIVKASVAALDGHCAALAAVAVERTG 64

50 Query: 78 RGVFEDKATKNLFACSHVNVNMRHRTKGVVIERDDVTGLTLIASPVGVVCGITPTNPTS 137
 RG+VEDKATKN+FACEHV + MRH KTVG+I D + G+T IARPVGVVCG+TP TNPTS
 Sbjct: 65 RGVFEDKATKNLFACSHVTHEMRHAKTVGLINDVPLGYTIEARPVGVVCGVTPVNTPS 124

55 Query: 138 TAIFESLSLSKTRNPIIFAFHPSAQESSAHARIVEDAAIAAGAPENCQWIRQSIDAT 197
 TAIFESLSLS+KTRNPI+V+FHPSA + S AA+IVEDAAIAAGAPENC+QWIR I+A+
 Sbjct: 125 TAIFESLSLSKTRNPIVFSFHPSALKCSIMAKIVEDAAIAAGAPENCQWIRGGIESAS 184

60 Query: 198 NALMHHDGIATIIATGGMAMVKAAYSOGKPALGVAGNVPYVYKESANTROAHDIVMSK 257
 N LNHG H+ATIIATGGMAMVKAAYS GKPALGVAGNVP Y+EK+ NI+QAA+D+VMSK

-344-

5
 10
 15
 20
 25
 30
 35
 40
 45

Sbjct: 185 NKLMMHFGVATILATGGNMMVKAYSSGKPALGVGAGNVPTIYIEKTCNIKQAANDVVMK 244
 Query: 258 SFDMGVCASEQAVIIDKEIYKFEVEEFSYHTYFVNKKKALLEEPCFGAKANSKNCAG 317
 SFDMGVCASEQAVIIDKEIY + VEE K + YF+N+IEKA LE+P FG A S +
 Sbjct: 245 SFDMGVCASEQAVIIDKEIYDQVVEEMKTLQAYFINSEKAKLEKFMFGVNAVSADVN 304
 Query: 318 AKLNPNIVCKSAWIAWIBQAGFTVPBGTNIIAARECTVSEKBLTREKLSFVIAVLKASST 377
 A+LNF G S W ADQ G VPE NI+ A C EV EBLTREKLSFV+A+LKAS+T
 Sbjct: 305 ARLNPKCPGMSQWFAFGVGIKVPEDCNICVACKEVGPNEPLTREKLSFVIAVLKASST 364
 Query: 378 EDGVKAKQMVFEENGLHSAIHTKADALAREPCTIRAIRVIVNNSPTPGTIGDVVNAF 437
 +DG+KA MVEENG GESAITH+ D + + +A R++ N+EG+ QGIG +YN
 Sbjct: 365 QGSDIKAEAMVEFMGRGHSAAIHSNDKAVVEKVALTMKACRILHAWTPSSQGGIGSTINWI 424
 Query: 438 LPSLTLGCGSYGRBSVGRNVAINLLMIKKVGRERNMNMFKVPSKTYFERDSTQYLQKC 497
 PS TLGCGSYG NSV N+ NLINIK++ RRHN+QMP+VP K +PE SI+YL +
 Sbjct: 425 WPSFTLGGSYGHSVANVTYHLLNINRGLADRRNLMQFRVPPKIPFESIRKYLAE 484
 Query: 498 RDUVRVIVTDHMFELGFLDRIIEQLDLRPNVYQIFARVEEPDSDITVMKGTDLMT 557
 +++ ++ IV+D M +LG+DR+++ L R N+V +IF +VEEDP I TV KG +M T
 Sbjct: 485 KELSKIPVSDRMYYKLGVDVDMVLKRRSNEVEIEIFIDVEPDSTQTVKGLAVMT 544
 Query: 558 FKEDTIALGGGSMDAQKVMWLYEQVEPDHDLVQKFMDIRKRAKFPFELGKKTFFVA 617
 F ED IIA+GGG MDAK+MKL YE PE DF + QKF+D+RKRAPFP +GKK + +
 Sbjct: 545 FGPDNIIAIGGGSMDAAKIMWLYEHPADFFAMKQKPIDLRKAPFPFTMGKKARLIC 604
 Query: 618 IPTSTGTGSEVTFPAVISDKANRKYPIADYSLTPTVAIVDPALVMTVPSTPAADTGMVD 677
 IPTSTGTGSEVTFPAVISD +KYP+ADYSLTP+VAIVDP M++P ADTG+DV
 Sbjct: 605 IPTSTGTGSEVTFPAVISDHETGKXYPLADYSLTPSVAIVDPMFTMSLPKRAIADTGLVD 664
 Query: 678 LTHATEAVYSQMANDYTDGLALQAIKIVFDYLSVKDADFEAREKMGNSATMGMAFAN 737
 L HATEAVYS MAM+YTDGLA +A+K+VF+ L +S + D EAREKMGNA+T+AGMAFAN
 Sbjct: 665 LVHATEAVYSVMAHETDGLARAVKLVFENLLKSY+NGLEAREKMGNAATLGAFAFAS 723
 Query: 738 AFLGISHSMAHKIGQAFTHVGRNTALILEYVIRYNTREAKTATWPKYNYRDEKQYD 797
 AFLG+ HSMHKG+GA FH HSR A+LLE+VIRYNG +P K A WPKYN+Y+AD+Y +
 Sbjct: 724 AFLGDHSMHKGMAFHLPHGRCAVLLFHVIRYNGQKPKLMMWPKYNYFKAQDQRYE 783
 Query: 798 IAKLLGLPAATFEAVESYAKAVYDLGIRLGKONFDDQIDSEKWEKESRELAFLAYED 857
 A++++GL TP E VE++AKA +L F+ IDE W K B+A LA+ED
 Sbjct: 784 LAQVGLKQNTFAGVEAFKAACEELMKATETITGFKANIDEAAMSKVPEALLAFED 843
 Query: 858 QCSFANFRLEWDMHMQEIIIDAYY 881
 QCSFANFR+PMW M++I++ AYY
 Sbjct: 844 QCSFANFRPMVIDMEKILCAAYY 867

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 875> which encodes the amino acid sequence <SEQ ID 876>. Analysis of this protein sequence reveals the following:

50
 55
 60
 65

Possible site: 55
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -3.66 Transmembrane 643 - 659 (642 - 660)
 INTEGRAL Likelihood = -1.81 Transmembrane 102 - 118 (102 - 118)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2466(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
 The protein has homology with the following sequences in the databases:
 >GP:AAA81906 GB:U04863 alcohol dehydrogenase 2 [Bntamoeba
 histolytica]
 Identities = 535/870 (61%), Positives = 669/870 (76%), Gaps = 3/870 (0%)
 Query: 6 NIVETTSVSVIIDALVQKGLAALSERMKLD--QBOVDYIVAKASVAAALDAHGEIAKHAYE 63

		+T +T +V I+ LV+K AL+E K + QE++DIV KASVAALD H LA A E	
	Subjct: 2	STQQTMTVDEHINQLVRKQVQVAKLEYLKPEYTGKIDYIVKKASVAALDQHCALAAAIVE	61
5	Query: 64	ETGRGVFEDKATKHLFACHEVNNMHQKTCVGLIERDDVTGLTLIARFVGVI CGITPTFN	123
		ETGRG+FDKATK++FACERV + MH KTVGLI D + G+T IARFVGVI+CG+TP FN	
	Subjct: 62	ETGRGI FDKATKNI FACHEVTHEMRKATVGLIINVDPLGYITRIARFVGVI+CGITPTFN	121
10	Query: 124	PTSTAIFKSLISIKTNPIIPAFHPGQSSGNAARTVRDAIYAGAPENCQW+ETPSL	183
		PTSTAIFKSLIS+KTNPI+P+FHPSA + S AA+TVRDAIYAGAPENCQW+E	
	Subjct: 122	PTSTAIFKSLISIKTNPIFVPSHPALSKCSIMAAKTVRDAIYAGAPENCQWIEFGI	181
15	Query: 184	EATNALMHGDIATILATGNNAMVKAAYSCKHALGVGRNVPAVVEKSNRQAAMHIV	243
		EAA+N LAMH G+ATILATGNNAMVKAAYS CKHALGVGRNVVP Y+EK+ HI+QAA+DV	
	Subjct: 182	EASNKLMHGHGATILATGNNAMVKAAYSSCKPALGVGRNVPTTYEKTCTIKQAANDUV	243
20	Query: 244	MSKSFDMGMVCASQNAVIIDKEIYDDFVAEPKSYHTYFVNKKEGALLERFCFGAKNSIN	303
		MSKSFDMGM+CASQNA IIDKEIYD V E K+ YF+N++EKA LE+P FG A+S	
	Subjct: 242	MSKSFDMGMVCASQNAVIIDKEIYDQVVEEMTGLGAYFINSEEGAKLEKPMFGVWYASD	301
25	Query: 364	ESREDSGVEKARQNVFENGHLSAAHITADASLAKFETGTRIRAIRVIMNSPSTPGGIGDYY	423
		E+ +DG++KA MVEFNG HSAATH+ D + +++ ++A R++ N+PS+ GGIG +Y	
	Subjct: 362	ENTODGIDKAEAMVEFNGHLSAAHSDKAVVEKYALTMACRLIHNTSSQGGISIV	421
30	Query: 424	NAPLSLTLGCGSYGRNAVDNVSADINLLNIKKVRRRNNMOWKVPKSPKTYFERDSIQYL	483
		N PS TLGCGSYG N+V NV+ NLLNIK++ RRRN+QWF+VP K +FE SI+YL	
	Subjct: 422	NYIMPSFTLGCSSYSGNSVSANVTYHLLNIKELADRNNLQWFRVPPKIFPEPSIRYL	481
35	Query: 484	QCKRDVERVMIVIDHAMVELGFLDIRIOLDLRNKVVOYFAEVEPDDPTTVMKSTEL	543
		+ +++ ++ IV+D M +LG+DR+++ L R N+V +IF +VEPDP I TV KG +	
	Subjct: 482	AEIKELSKIIFVSDRMWIKLGYVDRVMGLVGRNSKEVEIEIFIDVEPDPISIQTVQGLAV	541
40	Query: 544	MRTFKPTDIIALGGGSPMDAAKMWMLFYEOPEVDPHDLVOKFMDIRKRAFKFPELGKTK	603
		M TF PD IIA+GGGS MDAK+MWL YE PE DF + QKF+D+KRAFKFP +GKK +	
	Subjct: 542	MNTFGPNIIAIGGGSPMDAAKMWMLLYEHPEADFFAMKQKFDLEKRAFKFPTMOKKAR	601
45	Query: 604	FVAIPITSGTGSEVTPFAVISDKANNRKYPIADYSLTPTVAIVDPALVLTVPFGIAADTG	663
		+ IPTSGTGSEVTPFAVISD +KYP+ADYSLT+VAIVDP +++P ADTG	
	Subjct: 602	LICIPITSGTGSEVTPFAVISDHETGKKYPIADYSLTPTVAIVDPMTFSLPKRAIADTG	661
50	Query: 664	MDVLVHATEAYVSQMANPDTGLALQAIKIVFNLEKSVKTDAPREKQGNASTMAGMA	723
		+DVL HATEAYVS MAN++TDGLA +A+K+V+NL KS D EAREKQGNAT+AGMA	
	Subjct: 662	LDVLVHATEAYVSQMANPDTGLAREAVLKVFENLAKSY+NGLEAREKQGNASTMAGMA	720
55	Query: 724	PANAFGLISHSMHAKIGAQFHTVHGRTNAILLPVIRYNGTRPAKTATWPKYNYTRADEK	783
		PA+AFGL+ HSMHAK+GA FH HGR A+LLP+VIRYNG +P K A WPKYN+Y+AD++	
	Subjct: 721	PASAFGLMDHSMHAKVGAAFLPHGRCAVLLPHVIRYNGKPEKLANWPKYNYTKADQR	780
60	Query: 784	YQDIARKLGLPASTPREAVSYAKAVYDLCGRVQIMNQFKAQIDENSWKESRESLAYLA	843
		Y +A++GL +TP E VE++AKA +L FK IDE W E+A LA	
	Subjct: 781	YMKLAQMGVLCNTPEAGVEAPAKACEIMKATETITGFKKANIDEAANMSKVPEMALLA	840
65	Query: 844	YEDQCSANPRLEPMVDHMQEIRDAYGYGA 873	
		+EDQCSANPR+PMV N++I++ AYY A	
	Subjct: 841	FEDQCSANPRVPMVDHMKELKAAATYPIA 870	
66	An alignment of the GAS and GBS proteins is shown below:		
	Identities = 827/880 (93%), Positives = 852/880 (95%)		
70	Query: 12	NTEKTKAVETTDVALAIDTLVQGLKALDEMGRQLNQBOVDYIVAKASVAALDAPGSLALH	71
		NTE VETT V++ ID LVQ GL AL+EMR+L+QBOVDYIVAKASVAALDAPGSLA H	
75	Subjct: 1	NTECHNTVETTSVVTIDALVQGLKAALEEMRKLQBOVDYIVAKASVAALDAPGSLALH	60
80	Query: 72	AVETETGRGVFEDKATKNI FACHEVNNMRHTKTVGLIERDDVTGLTLIARFVGVI+CGITP	131

-346-

A EETGRGVFEDKATK+LFACEHVNNMRH KTVG+IREDDVGLTLIAKPVGV+CGITP
 Sbjct: 61 AYEETGRGVFEDKATKHLFACEHVNNMRHQKTVGI+IREDDVGLTLIAKPVGV+CGITP 120
 5 Query: 132 TTNPTSTAIKSLISLAKTRNP+IFAFHP+SAQESSAHAARTVRDA+IANGAPENCVQIBQ 191
 TTNPTSTAIKSLISLAKTRNP+IFAFHP+SAQESSAHAARTVRDA+IANGAPENCVQ+
 Sbjct: 121 TTNPTSTAIKSLISLAKTRNP+IFAFHP+SAQESSAHAARTVRDA+IANGAPENCVQVET 180
 Query: 192 PSIDATNALMNHDIATITLATGGNAMVKAAYS CGKPA LGVAGNVAYVEKSNIRQAAH 251
 PS++ATNALMNHDIATITLATGGNAMVKAAYS CGKPA LGVAGNVAYVEKSNIRQAAH
 10 Sbjct: 181 PSLEATNALMNHDIATITLATGGNAMVKAAYS CGKPA LGVAGNVAYVEKSNIRQAAH 240
 Query: 252 DIVMSKSPDNGMVCASQAVIIDKREIYK+VFPK+KREKALLKFCFGAKAN 311
 DIVMSKSPDNGMVCASQAVIIDKREIY +V FPKSYHTYFVNKREKALLKFCFGAKAN
 Sbjct: 241 DIVMSKSPDNGMVCASQAVIIDKREIYDDFVAFKSYHTYFVNKREKALLKFCFGAKAN 300
 15 Query: 312 SKNCAGAKLNPHIVGKSAVNIABQAGFTVPBGTNIIAABCTEVSEKEPLTREKLSPIVAV 371
 SKNCAGAKLNPHIVGK A NIABQAGFTVPBGTNIIAABCT EVSE EPLTREKLSPIVAV
 Sbjct: 301 SKNCAGAKLNPHIVGK PATNIABQAGFTVPBGTNIIAABCTEVSENEPLTREKLSPIVAV 360
 20 Query: 372 LKASSTEDGVEKARQMFVFNGLGHSAAIHTADDLAREFGTRIRAIRVWNSPSTFGGIG 431
 LK+ES EDGVEKARQMFVFNGLGHSAAIHT DA+LA+EFGTRIRAIRVWNSPSTFGGIG
 Sbjct: 361 LKESREDDGVEKARQMFVFNGLGHSAAIHTADAEKAEFGTRIRAIRVWNSPSTFGGIG 420
 25 Query: 432 DVYNAFLPSLTLCGSGYGRNVGDSAINLLNKKVGRRRNNMQFKVPSKTYFERDSI 491
 DVYNAFLPSLTLCGSGYGRN+VGDVSAINLLNKKVGRRRNNMQFKVPSKTYFERDSI
 Sbjct: 421 DVYNAFLPSLTLCGSGYGRNVGDSAINLLNKKVGRRRNNMQFKVPSKTYFERDSI 480
 30 Query: 492 QYLQKCRDVERVMIVTDHAWVELGFLDRIIBOLDLRNRKVYQIFASVEPDDITTVHGK 551
 QYLQKCRDVERVMIVTDHAWVELGFLDRIIBOLDLRNRKVYQIFASVEPDDITTVHGK
 Sbjct: 481 QYLQKCRDVERVMIVTDHAWVELGFLDRIIBOLDLRNRKVYQIFASVEPDDITTVHGK 540
 Query: 552 TDLNRFTFKPDTIIALGCGSPMDAAKVMWLFYEQEVDVFDLVQKFMDIRKRAKFPPELGK 611
 T+LNRFTFKPDTIIALGCGSPMDAAKVMWLFYEQEVDVFDLVQKFMDIRKRAKFPPELGK
 Sbjct: 541 TELNRFTFKPDTIIALGCGSPMDAAKVMWLFYEQEVDVFDLVQKFMDIRKRAKFPPELGK 600
 35 Query: 612 KTKFVAIPTTSCTGSEVTPPAVISDKANNRKYPIDYSLTPTVAIVDPALVTVPGFIAA 671
 KTKFVAIPTTSCTGSEVTPPAVISDKANNRKYPIDYSLTPTVAIVDPALVTVPGFIAA
 Sbjct: 601 KTKFVAIPTTSCTGSEVTPPAVISDKANNRKYPIDYSLTPTVAIVDPALVTVPGFIAA 660
 40 Query: 672 DTGMDVLTHATEAVVSQMANDYTDGLALQAKIVFDYLSRVKIDAFERKEMENASTMA 731
 DTGMDVLTHATEAVVSQMAND+TDGLALQAKIVFD LK+SVK ADFERKEMENASTMA
 Sbjct: 661 DTGMDVLTHATEAVVSQMANDPTDGLALQAKIVFDNLEKSVKIDAFERKEMENASTMA 720
 45 Query: 732 GMAFANAPLGISHMAHKIGAQPHTVHGRTNAILLPVIRYNGTRPAKTATWPKYNYRA 791
 GMAFANAPLGISHMAHKIGAQPHTVHGRTNAILLPVIRYNGTRPAKTATWPKYNYRA
 Sbjct: 721 GMAFANAPLGISHMAHKIGAQPHTVHGRTNAILLPVIRYNGTRPAKTATWPKYNYRA 780
 Query: 792 DEKYQDIKLLGLPASTPERAVESYAKAVYDLGCRVGIQNMFKQIDENEMKESRELA 851
 DEKYQDIKLLGLPASTPERAVESYAKAVYDLG R+GI+MNF+ QSIDENEMKESRELA
 50 Sbjct: 781 DEKYQDIKLLGLPASTPERAVESYAKAVYDLGCRVGIQNMFKQIDENEMKESRELA 840
 Query: 852 FLAYEDQCSANPRLFWNDHMQEIIEDAYGYAERPERK 891
 +LAYEDQCSANPRLFWNDHMQEIIEDAYGY ERPERK
 Sbjct: 841 FLAYEDQCSANPRLFWNDHMQEIIEDAYGYAERPERK 880
 55

A related GBS gene <SEQ ID 8533> and protein <SEQ ID 8534> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10
 Mcg: Discrim Score: -4.68
 60 GvH: Signal Score (-7.5): -2.48
 Possible site: 21
 >>> Seems to have no N-terminal signal sequence
 ALOM program count: 1 value: -2.66 threshold: 0.0
 INTEGRAL Likelihood = -2.66 Transmembrane 100 - 116 (99 - 117)
 65 PERIPHERAL Likelihood = 3.61 173
 modified ALOM score: 1.03

*** Reasoning Step: 3

----- Final Results -----

5 bacterial membrane --- Certainty=0.2062 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

SEQ ID 8534 (GBS432) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 5; MW 66kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 7; MW 41kDa).

GBS432-GST was purified as shown in Figure 223, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

15 Example 274

A DNA sequence (GBSx0299) was identified in *S.agalactiae* <SEQ ID 877> which encodes the amino acid sequence <SEQ ID 878>. Analysis of this protein sequence reveals the following:

Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.3444 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
25 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 880.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

30 Example 275

A DNA sequence (GBSx0300) was identified in *S.agalactiae* <SEQ ID 881> which encodes the amino acid sequence <SEQ ID 882>. Analysis of this protein sequence reveals the following:

Possible site: 26
>>> Seems to have a cleavable N-term signal seq.
35 INTEGRAL Likelihood = -8.39 Transmembrane 74 - 90 (69 - 94)
 INTEGRAL Likelihood = -5.31 Transmembrane 168 - 184 (163 - 186)
 INTEGRAL Likelihood = -4.83 Transmembrane 34 - 50 (29 - 52)
 INTEGRAL Likelihood = -0.75 Transmembrane 202 - 218 (202 - 219)

40 ----- Final Results -----

 bacterial membrane --- Certainty=0.4354 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA17305 GB:AL021926 hypothetical protein RV0111 [Mycobacterium
 tuberculosis]
 Identities = 70/218 (32%), Positives = 104/218 (47%), Gaps = 12/218 (5%)

50 Query: 9 VRITGLLLVLLYHFFKNSFPGGFVGVDDIFFTFSGPLITALLIDEFSKYKIDFVSFCRRR 68
 +R + LVL H GGF+GVD FF SGFLIT+LL+DE +T +ID F RR

-348-

Sbjct: 39 LRAIAVALVLASHGGIPQMGGGFVGVDAPFVLSGLITLILDLGRIGRIDLSGFWRIR 98

Query: 69 FYRIFFPLVLMVLVTIPFVFLVKSDFRASIGSQIMTAIGFTSNFYELTGGNYESQFI-P 127
R+ P LVIMVL L + S + A +T+N+ + +Y +Q P

5 Sbjct: 99 ARRLLEALVLMVLTVSARALFFDQALTLGRSDAIAAFIATANWRFFVAGNTIDYFTQGAFF 158

Query: 128 HLPVHTWLSLIEVHFVLMWGL----TVWLLSKRSKQKQLRGTLFLISMGIQVGSFLTMF 183
HTWSL +E +YV+W L LL+ R++ ++ R T+ + F ++ L

10 Sbjct: 159 SPLQHTWLSLGVBEQYVYVWPLLIGATLLAARAR-RRCRRATVGVGFAPFLIASLSTM 217

Query: 184 VRAPFVDNFST-----IYFSTLSHIPFFFLGAMVATI 215
A F++ IYF T + +G+ A +

Sbjct: 218 ASATAVAFTSAATRDRIYFGDITRAQALLTGSAAAL 255

- 15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 879> which encodes the amino acid sequence <SEQ ID 880>. Analysis of this protein sequence reveals the following:

Possible site: 46
>>> Seems to have an uncleavable N-term signal seq

20	INTEGRAL	Likelihood = -10.83	Transmembrane	325 - 341 (313 - 346)
	INTEGRAL	Likelihood = -9.29	Transmembrane	237 - 253 (234 - 258)
	INTEGRAL	Likelihood = -7.91	Transmembrane	166 - 182 (162 - 188)
	INTEGRAL	Likelihood = -6.10	Transmembrane	72 - 88 (68 - 92)
	INTEGRAL	Likelihood = -4.09	Transmembrane	264 - 280 (260 - 281)
	INTEGRAL	Likelihood = -2.87	Transmembrane	371 - 387 (370 - 390)
25	INTEGRAL	Likelihood = -2.66	Transmembrane	34 - 50 (32 - 50)
	INTEGRAL	Likelihood = -1.91	Transmembrane	3 - 19 (3 - 15)
	INTEGRAL	Likelihood = -0.85	Transmembrane	136 - 152 (136 - 154)

----- Final Results -----

30 bacterial membrane --- Certainty=0.5331(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

35 Identities = 167/226 (73%), Positives = 195/226 (85%)

Query: 1 MRIKWPSLVIRITGLLLVLLYHFFKNSFFGGFVGVDIFFFTFSGLITALLIDERSKTKCID 60
MRIKWPS VR+TGLLLVLLYHFFKN FFGGF+GVDIFFFTFSGL+LITALLIDE++K + ID

40 Sbjct: 1 MRIKWPSFVRVTGLLLVLLYHFFKNFVFGGFVGVDIFFFTFSGLITALLIDEYTKKESID 60

Query: 61 FVSPCRRRFYRIFFPLVLMVLVTIPFVFLVKSDFRASIGSQIMTAIGFTSNFYELTGGN 120
+ F +RRFYRI FPLVLM+L+TIPF FL+K DF A+IGSQI LGFTN YELITG +

Sbjct: 61 IIGFLKRKFYRIVPPLVLMILLTIPPTFLIKKDFIANIGSQITAVLGFTTNIYEILTGSS 120

45 Query: 121 YESQFIPHLFVHTWLSLIEVHFVLMWGLTVWLLSKRSKQKQLRGTLFLISMGIQVGSFL 180
YESQFIPHLFVHTWLSL+IEVHFT+ WG+ VWLL++R + QKQLRG LFLIS+GIF +SPL

Sbjct: 121 YESQFIPHLFVHTWLSLIEVHFVLMWGVFVWLLARRKRTQKQLRGLLFLISLGI FAISPL 180

50 Query: 181 TMPVRAPFVDNFSTIYFSTLSHIPFFFLGAMVATIGIRETGRFK 226
+MF-R+F NFS IYFS+LSH FPFLLGAM ATI+GI E T RE+

Sbjct: 181 SMFIRSFMTSNFSLSIYFSSLSHSFPFLGAMFATTIGINETVRPQ 226

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

55 Example 276

A DNA sequence (GBSx0302) was identified in *S.agalactiae* <SEQ ID 883> which encodes the amino acid sequence <SEQ ID 884>. Analysis of this protein sequence reveals the following:

Possible site: 38
>>> Seems to have a cleavable N-term signal seq.

60

----- Final Results -----

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database:

- 1GB:AE004818 hypothetical protein [Pseudomonas aerug...]
 1GB:AE004818 hypothetical protein [Pseudomonas aerug...]
 1GB:AE004818 hypothetical protein [Pseudomonas aerug...]
 1GB:AE004818 hypothetical protein [Pseudomonas aerug...]
 1GB:AE004818 hypothetical protein [Pseudomonas aerug...]
- >GP:AG07403 GB:AE004818 hypothetical protein [Pseudomonas aeruginosa]
 Identities = 33/80 (41%), Positives = 50/80 (62%)
- 15 Query: 45 KIVGSI VNHMTGKGLTYENGDIYKGFVNGVFBGKGTFFSVHGMSTGDFPKKGQPDGQ 104
 +Y G +V+ + G+G+L Y+NG +Y G F +G+ G GT+ G Y+G F G DQQ
 Sbjct: 39 RYRGELVDGRLEQQRLDYDNGAWYAGRFEHGLLHGHTWQADSSRYSGGFAAGLFDGQ 98
- 20 Query: 105 GRLENAHQGVYKGTFFKGIY 124
 GRL + VY+G F+QG+
 Sbjct: 99 GRLENAHQGVYKGTFFKGIY 118
 Identities = 31/91 (34%), Positives = 46/91 (50%), Gaps = 2/91 (2%)
- 25 Query: 34 QGVFSYDGGKIKYVGSIVNHMTGKGLTYENGDIYKGFVNGVFBGKGTFFSVHGMSTG 93
 QG YD G Y G + + G+G +G Y G F G G+G+G G Y
 Sbjct: 52 QGRLDYDNGAW-YAGRFEHGLLHGHTWQADSSRYSGGFAAGLFDGQGRLENAHQGVY 110
- 30 Query: 94 GDFPKKGQPDGGRLENAHQGVYKGTFFKGIY 124
 G F+G DG+G L + + Y+G F+G+Y
 Sbjct: 111 GDFPKKGQPDGGRLENAHQGVYKGTFFKGIY 140
 Identities = 31/91 (34%), Positives = 42/91 (46%), Gaps = 1/91 (1%)
- 35 Query: 32 SSQGVFSYDGGKIKYVGSIVNHMTGKGLTYENGDIYKGFVNGVFBGKGTFFSVHGMSTG 91
 S QG G +Y GS + G+G + G+ Y G F +G GKG + G
 Sbjct: 141 SSQGVFSYDGGKIKYVGSIVNHMTGKGLTYENGDIYKGFVNGVFBGKGTFFSVHGMSTG 99
- 40 Query: 92 YTGDFPKKGQPDGGRLENAHQGVYKGTFFKGIY 122
 Y G FK Q QGGR + + V+ G F +G
 Sbjct: 200 YTGDFPKKGQPDGGRLENAHQGVYKGTFFKGIY 230
 Identities = 31/91 (34%), Positives = 45/91 (49%), Gaps = 4/91 (4%)
- 45 Query: 34 QGVFSYDGGK----IKYVGSIVNHMTGKGLTYENGDIYKGFVNGVFBGKGTFFSVHGMSTG 89
 QG+F +G +Y G +G+G L +G Y+G F G G+G+G F G
 Sbjct: 115 QGDFLEGGSLLEQQOTRYRGGRFKGLYSQGGTLDGDSGRYQGSFRQGRLENAHQGVY 174
- Query: 90 WSYTGDFPKKGQPDGGRLENAHQGVYKGTFFKGIY 120
 Y G F+ QG +G+G + + Y G FK
 Sbjct: 175 WSYTGDFPKKGQPDGGRLENAHQGVYKGTFFKGIY 205
 Identities = 26/87 (32%), Positives = 45/87 (51%), Gaps = 1/87 (1%)
- 50 Query: 34 QGVFSYDGGKIKYVGSIVNHMTGKGLTYENGDIYKGFVNGVFBGKGTFFSVHGMSTG 93
 +G FS G +Y G+ + + GKG+ + +GD Y G F + F G+G + S G +
 Sbjct: 166 GGSFSDSQN-QYAGTFRDGLNGKGRNSGPDGDRYVQGFQKQPHQGRYGSASGVDWI 224
- 55 Query: 94 GDFPKKGQPDGGRLENAHQGVYKGTFFKGIY 120
 G F +G +G G L + Y+G F+
 Sbjct: 225 GDFPKKGQPDGGRLENAHQGVYKGTFFKGIY 251
 Identities = 26/89 (31%), Positives = 43/89 (47%), Gaps = 2/89 (2%)
- 60 Query: 34 QGVFSYDGGKIKYVGSIVNHMTGKGLTYENGDIYKGFVNGVFBGKGTFFSVHGMSTG 93
 QG + G + Y G G+G L + G Y+G F G+ +G+GT G Y
 Sbjct: 98 QGRLAMADGS-VYGGFRQGLFDGEGSLR-QGQTRYRGGRFKGLYSQGGTLDGDSGRYQ 155
- 65 Query: 94 GDFPKKGQPDGGRLENAHQGVYKGTFFKGIY 122
 G F+G+G +G+G + Y GTF+ G
 Sbjct: 156 GDFPKKGQPDGGRLENAHQGVYKGTFFKGIY 184

-350-

Identities = 25/80 (31%), Positives = 37/80 (46%)

Query: 45 KYVGSIVNHHMTGKGLTYBNKDYKQDFVNGVFBGKGTFFSVHGWSTYTGDFKKQGPQ 104
 +YVG ++ G+G+ +GD + G F G G G + G Y G F+ + GQ
 Sbjct: 199 RYVVGFKINQFHGQGRYESAGSDWIGRFSBGALNGPGLLQADQSRYSRGQGPQWRPHQ 258

Query: 105 GRINAKNKIKYKGTFFKQGIY 124
 G L + Y+G F G Y
 Sbjct: 259 GLLGLDGTTRYBGSFAAGAY 278

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 885> which encodes the amino acid sequence <SEQ ID 886>. Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -13.16 Transmembrane 20 - 36 (12 - 41)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.6265(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAAL6606 GB:D90899 hypothetical protein [Synecocystis sp.]
 Identities = 37/89 (41%), Positives = 49/89 (54%), Gaps = 6/89 (6%)
 Query: 48 KGRMHTT-----GVVNHKMNBSGKLVYPNGDIYEGTFKDLFBGKGTFFAKTGWLYNG 101
 KG YT G V+ ++NG GK Y NGD YEGT K+G +G+G F G Y G
 Sbjct: 141 KGTFTYINGDRCSGTIVVQGEINSGKCEINNGDQYBGTLENGQPDGEGIFRPAAGSGSYEG 200
 Query: 102 EFHKGQANGKGLVAKANNKKYKGIKQGI 130
 EF G+ +G+G N ++G FKQ+
 Sbjct: 201 EFQSGEFSGGQTRIFANGNRPQSQFQGL 229

An alignment of the GAS and GBS proteins is shown below:

Identities = 68/126 (53%), Positives = 93/126 (72%)

Query: 1 MGNFKITRTHLEILSLIIIVFGLSVFTLTSSQGVFSYDGGKIKYVGSIVNHHMTGK 60
 +K + ITR LEI+S+I+V +SVF++ S++ +YD G++ Y G ++NH M G+GK
 Sbjct: 8 VKKMSITRAKLEIVSVIVILVCAISVFSVRLSNKTSLTVDKGRMHYTGYYVNHQNGBEK 67
 Query: 61 LTYENGDDYKQDFVNGVFBGKGTFFSVHGWSTYTGDFKKQGPQGRINAKNKIKYKGTFF 120
 L Y NGD Y+G F +G+FBGKGTFF + GW Y G+F KGQ +G+G L AKN KUYKG FK
 Sbjct: 68 LVYPNGDIYBGTFFKDLFBGKGTFFAKTGWLYNGEFHKGQANGKGLVAKANNKKYKGIK 127
 Query: 121 QGIYQK 126
 QGI+QK
 Sbjct: 128 QGIYQK 133

SEQ ID 884 (GBS139) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 19 (lane 3; MW 13kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 22 (lane 2; MW 38.2kDa), in Figure 24 (lane 7; MW 38kDa) and in Figure 33 (lane 7; MW 38.2kDa).

The GBS139-GST fusion product was purified (Figure 200, lane 2) and used to immunise mice. The resulting antiserum was used for FACS (Figure 287), which confirmed that the protein is immunoaccessible on GBS bacteria.

-351-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 277

A DNA sequence (GBSx0303) was identified in *S. agalactiae* <SEQ ID 887> which encodes the amino acid sequence <SEQ ID 888>. This protein is predicted to be holliday junction dna helicase *ruvB* (*ruvB*). Analysis of this protein sequence reveals the following:

Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4386 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CA875331 GB:Y15896 RuvB protein [Bacillus subtilis]
Identities = 196/322 (60%), Positives = 254/322 (78%)

Query: 3 RFLDSNAGDESELVETLRPQYLREYIGQDKVKQQLKPIEAAKLDESLDHVLLPGPPG 62
R + S+A E ++E++LRPQ L +YIGQ KVK+ L++FI+AAK+R E+LORVLL+GPPG
Sbjct: 4 RLVSSDNNHSEVIBQGLFQNLAQYIGQHKVENLRVFIADAMNRQETLDHVLLYGPFG 63

Query: 63 LGKTTAFVIANELGVLEKQTSGPAIEKSGDLVAIALDLPGDVLFDIEIHRMPMAVEEV 122
LGKTT+A ++ANE+GV L+ TSGPATB+ GDL AIL LEPGDVLFDIEIHR+ ++EEV
Sbjct: 64 LGKTTIASIVANEMGVLEKRTTSGPATERPGLAAILTALPGDVLFDIEIHLRHSIEEV 123

Query: 123 LYSAMEDFYIDIMIGAGETSRVHLDPFPFTLIGMTTRAGMLSNPLARFQITGHMEIYE 182
LY AMEDF +DI+IG G ++REV LDLPFPFTL+GNTTR G+L+ FLR RFG+ +EYI
Sbjct: 124 LYPAMEDFCLDIVIGNGPEAKSRVLDLPFPFTLVGATTRVGLLTAPLADRFQVMRSLEYIT 183

Query: 183 ENDLTIIEKTADIFRMKITVRAASELAKRSRGTPRIANRLKRVRYTAQIMSGDLIDN 242
+ +L +I+ RTAD+FE++I A B+ARRSRGTPR+ANELL+RVRD+AQ++GD I ++
Sbjct: 184 QEELADIVTKTADVFEVEIDKPSALSIARRSRGTPKRVANRLKRVRDFAQVLGDSERITD 243

Query: 243 ITDKALTHLDVHEGLDYVDKILRTWIMBNYGGPVLGLTSLVNTASSRDTVDNMYEPTL 302
I+ AL L VD GLD+D K+L MIE +NGSPVGL T+S I SE T+ED+YEPTL
Sbjct: 244 ISQALERLQVDRDLGLDHDHKLKANGIEKFNQGPVGLDTISNTIGESHTIEDVYEPTL 303

Query: 303 IQGSPIMRTTRGRVATVKAEH 324
+Q GPI RT GR+ T Y H
Sbjct: 304 LQIQPIQRTFGRIVTPAVVHH 325

A related GBS nucleic acid sequence <SEQ ID 10943> which encodes amino acid sequence <SEQ ID 10944> was also identified.

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 889> which encodes the amino acid sequence <SEQ ID 890>. Analysis of this protein sequence reveals the following:

Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0686 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 282/327 (86%), Positives = 306/327 (93%)

-352-

- Query: 1 MTRFLDSAMGDEBLVRETLRPQYLKRYIGQDKVKDQLKIFTEAKLRDESLDHLVLP 60
M R LD++ MG+EE +RTLRPQYL EYIQQDKVK+Q IFTEAAK RDESLDHLVLP 60
Sbjct: 25 MARILNDINVMGRNFSRDLRPQYLHXYIQQDKVKEQFAIFTEAAKRRDESLDHLVLP 84
- 5 Query: 61 PGLGKTMTAFVIANELGVNLKQTS GPAI EKS GDLVAI LNDLEPGDVL FIDEIHRMPVAE 120
PGLGKTMTAFVIANELGVNLKQTS GPAI EK+GDLVAI LN+LEPGD+L FIDEIHRMPM+VE 120
Sbjct: 85 PGLGKTMTAFVIANELGVNLKQTS GPAVKEAGDLVAI LNELEPGDIL FIDEIHRMP+SVK 144
- 10 Query: 121 EVLVSAMSDFYIDIMIGAGETSRSHLLDPPPTLIGATTAGMLSNPLARARFGITGMSY 180
EVLVSAMSDFYIDIMIGAG+TSRS+HLLDPPPTLIGATTAGMLSNPLARARFGITGMSY 180
Sbjct: 145 EVLVSAMSDFYIDIMIGAGETSRSHLLDPPPTLIGATTAGMLSNPLARARFGITGMSY 204
- 15 Query: 181 YEENDLTIETITADIFEMKITYEAASELARRSRGTPRIANRLKRVRDYQAQIMGDSLID 240
Y+E DLTEI+EKTA IPE+KI +EAA +LA RSRGTPRIANRLKRVRDYQAQI+GDS+I 240
Sbjct: 205 YQERDLTEI VERTATTIPEIKIDHEAARKLACRSRGTPRIANRLKRVRDYQAQIIGDSIT 264
- 20 Query: 241 DNTIDKALTMLDVDESGLDYVDKILKRTMIEMYGSGFVGLGTLGVNIAERDVTEDMTEP 300
ITD+ALTMLDVD SGLDY+DQKILKRTMIEM YGSGFVGLGTLGVNIAER+TVE+MYEP 300
Sbjct: 265 AQITDRALTMLDVDESGLDYIDKILKRTMIEM YQGSGFVGLGTLGVNIAERNTVEEMTEP 324
- Query: 301 YLIGNGFIMRTITGRVATVKAYEHLGY 327
YLIGNGF+MRTITGRVAT KAY HLGY 327
Sbjct: 325 YLIGNGFIMRTITGRVATVKAYEHLGY 351
- 25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 278

A DNA sequence (GBSx0304) was identified in *S.galactiae* <SEQ ID 891> which encodes the amino acid sequence <SEQ ID 892>. Analysis of this protein sequence reveals the following:

- 30 Possible site: 43
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -2.87 Transmembrane 157 - 173 (157 - 174)
INTEGRAL Likelihood = -1.49 Transmembrane 205 - 221 (205 - 222)
- 35 ----- Final Results -----
bacterial membrane --- Certainty=0.2147(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 40 The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 893> which encodes the amino acid sequence <SEQ ID 894>. Analysis of this protein sequence reveals the following:

- possible site: 56
>>> Seems to have no N-terminal signal sequence
- 45 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3097(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
- 50

An alignment of the GAS and GBS proteins is shown below:

- Identities = 130/303 (42%), Positives = 202/303 (65%)
- 55 Query: 1 MLGHFGSKVRNLRVTRNITREDFGQDETSLGVQLARIEGQSIPNLTYAHYIAQLNVK 60
ML+HFG KV+ LR+ I+RED QGDE+ELGVQLARIE GQSIP+L+K +IAK LNV 60
Sbjct: 1 MLGHFGSKVRNLRLEKRISREDFGQDETSLGVQLARIEQGQSIPNLTKSVITIAKLVN 60
- Query: 61 LDILTGGESLELPKRYKELKYLILRIPTTADARLKLRECPDHIPEEPYDNLPEECLEA 120
+ LT G LELPKRYKELKYLILR PTY D +L++RQ QPD IPE+YD LPE+E + 120

-353-

Sbjct: 61 VGYLTGADLELPKRYKLYLILRTPTTMDGKIQVREBQDFRIFREDYYKLPSEKII 120
 Query: 121 IDSLQAKFEVYQTGSDINPGVEVLCBDFKVKYKKTLDNLIIDLF/TCVVSVKFNRA 180
 ID LQA + + + NPG+++L E P+++K K ++ NDLI+++L+L + + +
 5 Sbjct: 121 IDCLQATLDLTLSENTNPGIDLAQYFQIKTKVFRQNDLILLELYLAYLDIEBQGY 180
 Query: 181 FTEKEVQFTICTKLISQNHKLTAEDLFWFMEVLNCFVGLCLNSECLAEMLAEVSRQTMV 240
 K + ++ L Q + ++LF N +++ + L N + L + +E+S+ + M
 10 Sbjct: 181 SKKIFDSELDNLSEQFQFELDELFIWNKIIDISLSLKNKRLDMLKAEIMSQKINA 240
 Query: 241 STHDFHMKPLYPFYQKQYF ITIDNDIKSAENAYQGSIMFSKIMDDKHLIKKLELWQREDI 300
 D+++MP+ + +WKYF+ DI AE ++ + +E+M D+L KL EW+D+
 Sbjct: 241 KIQDWRMPILKLIENKYFLIKQKDIKABQSFNKAELFAGMTADQYLENKLIQEWEKDV 300
 15 Query: 301 TGH 303
 +
 Sbjct: 301 KSY 303

SEQ ID 892 (GBS319) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell
 20 extract is shown in Figure 40 (lane 4; MW 37kDa). It was also expressed in *E. coli* as a GST-fusion
 product. SDS-PAGE analysis of total cell extract is shown in Figure 46 (lane 7; MW 62kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 279

25 A DNA sequence (GBSx0305) was identified in *S. galactiae* <SEQ ID 895> which encodes the amino acid
 sequence <SEQ ID 896>. This protein is predicted to be adenylosuccinate lyase (purB). Analysis of this
 protein sequence reveals the following:

Possible site: 35
 >>> Seems to have no N-terminal signal sequence
 30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3358(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 35

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04344 GB:AP001509 adenylosuccinate lyase [Bacillus halodurans]
 Identities = 326/430 (75%), Positives = 366/430 (84%)
 40 Query: 1 MIERYSRPFEMAIWTSEKNRYAMLEVEILADEAMAELEIPKEVDKAKIREKADFDIRIL 60
 MIERY+RPFM AIWTEEN+Y+AMLEVEI+A ENMAELGEIPKEVD KIRE A FD++RIL
 Sbjct: 1 MIERYTRPFMGAIWTEENRYQAMLEVEIACENMAELGEIPKEVDKIKREHASPFVERIL 60
 45 Query: 61 EIRQDTRHDVVAFTFVAVTSVETLGEERKWHYGLTSTDVVDTAAGGLYKQANDIIRDLNF 120
 EIRQ+TRHDVVAFTFVAVTSVETLGEERKWHYGLTSTDVVDTA YL KQAN+II DL F
 Sbjct: 61 EIRQDTRHDVVAFTFVAVTSVETLGEERKWHYGLTSTDVVDTALSGLKQANEIIRADLVRF 120
 50 Query: 121 TNIVADKAKCHKFTIMKRGTHGVHARPTTFGLKLAITWSEMKRNIREFEHAAGVAVGKI 180
 +I+ +KA EHK+T+MKGTHGVHARPTTFGLKLAITWSEMKRNLRRFRILAASTVRVGLK 180
 Sbjct: 121 LDTLKEKALEHKYTYMKGTHGVHARPTTFGLKLAITWSEMKRNLRRFRILAASTVRVGLK 180
 55 Query: 181 SGAVGNFANIPPFVEQYVCDKLGIRPQIRISTQVLRDILHARYFAVLASIAITSIERMATKI 240
 SGAVG +ANI PFVEQYV+C+KLG+ LSTQ L RD HARY A LA IATSIIE+ A EI
 Sbjct: 181 SGAVGTIYANIDPFVEQYVCEKLGIRAPISITQLRDRHARYMATIALIATSTIEKFAVKI 240
 Query: 241 RGLQKSETRVEEFPFAKQKQKSSAMPHEKNDIGSENNITGLARVIRGMBVAYENVALWHE 300
 RGLQKSE REVEE+FAKQKQKSSAMPHEKNDIGSENNITG+ARV+RGMB+ AYEN LWEH
 Sbjct: 241 RGLQKSETRVEEFPFAKQKQKSSAMPHEKNDIGSENNITGIARVVRGMLAAYENVPLWHE 300

-354-

Query: 301 RDISHSSAERIITPDTTILIDYMLNRFGNIVKNIIVFPENMMKRMESTFGLIYSQRVWLK 360
 RDISHSSAERIIT PD TI I+YMLNRFGNIVKNIIVFPEN RNM T+GLIYSQRV+L
 Sbjct: 301 RDISHSSAERIITPDATIAINYMLNRFGNIVKNIIVFPENMKRMKRTYGLIYSQRVLS 360

5 Query: 361 LIEKGMTREAYDLVOPKTAYSNDNQVDFKPLLEDTKVTSCLTQREIDELFNPIYYTKR 420
 LI+KGM RREAYDLVOPK +W+ V + L+E++ ++TS L+ EEI+ F+ ++ K
 Sbjct: 361 LIDKGMVREAYDLVQPKAMEAMEKGVQFRLEVEQREIRITSVLSPRETRACFDYNHHLKH 420

10 Query: 421 VDDIFERLGL 430
 VD IFERLGL
 Sbjct: 421 VDTIFERLGL 430

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 897> which encodes the amino acid sequence <SEQ ID 898>. Analysis of this protein sequence reveals the following:

15 Possible site: 35
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3358 (Affirmative) < succ>
 20 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 422/430 (98%), Positives = 428/430 (99%)

25 Query: 1 MIERYSRPEMAIWTENKYNAMLEVFRIILADEAWAELGEIPKRDVAKIREKADFDIDRL 60
 M+ERYSRPEMAIWTENKYNAMLEVFRIILADEAWAELGEIPKRDVAKIREKADFDIDRL
 Sbjct: 1 MIERYSRPEMAIWTENKYNAMLEVFRIILADEAWAELGEIPKRDVAKIREKADFDIDRL 60

30 Query: 61 EIEQDTRHDVAVFTRAVSETLGEERKOVVHYGLTSDVVLTAGVLYKQANDIIRDLNF 120
 EIEQDTRHDVAVFTRAVSETLGEERKOVVHYGLTSDVVLTAGVLYKQANDIIRDLNF
 Sbjct: 61 EIEQDTRHDVAVFTRAVSETLGEERKOVVHYGLTSDVVLTAGVLYKQANDIIRDLNF 120

35 Query: 121 TNIVADKAKHEKPTIMMGRTHGVHAEPTTFGLKLTWYSEMKRNIERPEHAAAGVERAGKI 180
 TNIVADK+EHK TIMMGRTHGVHAEPTTFGLKLTWYSEMKRNIERPEHAAAGVERAGKI
 Sbjct: 121 TNIVADKAKHEKPTIMMGRTHGVHAEPTTFGLKLTWYSEMKRNIERPEHAAAGVERAGKI 180

Query: 181 SGAVGNFANIPFPVBYVCCKLGRPQESTQVLPRDLHAIFYAVLASIATSIERMATEI 240
 SGAVGNFANIPFPVBYVCCKLGRPQESTQVLPRDLHAIFYAVLASIATSIERMATEI
 40 Sbjct: 181 SGAVGNFANIPFPVBYVCCKLGRPQESTQVLPRDLHAIFYAVLASIATSIERMATEI 240

Query: 241 RGLQKSEQRVEEFPAGQKQGSAMPKRNPIGSENTGLARVIRGHMVTAYENVALHE 300
 RGLQKSEQRVEEFPAGQKQGSAMPKRNPIGSENTGLARVIRGHMVTAYENVALHE
 45 Sbjct: 241 RGLQKSEQRVEEFPAGQKQGSAMPKRNPIGSENTGLARVIRGHMVTAYENVALHE 300

Query: 301 RDISHSSAERIITPDTTILIDYMLNRFGNIVKNIIVFPENMMKRMESTFGLIYSQRVWLK 360
 RDISHSSAERIITPDTTILIDYMLNRFGNIVKNIIVFPENMMKRMESTFGLIYSQRVWLK
 Sbjct: 301 RDISHSSAERIITPDTTILIDYMLNRFGNIVKNIIVFPENMMKRMESTFGLIYSQRVWLK 360

50 Query: 361 LIEKGMTREAYDLVOPKTAYSNDNQVDFKPLLEDTKVTSCLTQREIDELFNPIYYTKR 420
 LIEKGMTREAYDLVOPKTAYSNDNQVDFKPLLEDTKVTSCLTQREIDELFNPIYYTKR
 Sbjct: 361 LIEKGMTREAYDLVOPKTAYSNDNQVDFKPLLEDTKVTSCLTQREIDELFNPIYYTKR 420

55 Query: 421 VDDIFERLGL 430
 VDDIF+RLG+
 Sbjct: 421 VDDIFKRLGI 430

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-355-

Example 280

A DNA sequence (GBSx0306) was identified in *S.agalactiae* <SEQ ID 899> which encodes the amino acid sequence <SEQ ID 900>. Analysis of this protein sequence reveals the following:

```

Possible site: 45
5  >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -16.24    Transmembrane    145 - 161 ( 119 - 167)
    INTEGRAL    Likelihood = -9.98     Transmembrane    125 - 141 ( 119 - 144)
    INTEGRAL    Likelihood = -9.29     Transmembrane    28 - 44 ( 23 - 51)
10  INTEGRAL    Likelihood = -7.01     Transmembrane    196 - 212 ( 193 - 220)
    INTEGRAL    Likelihood = -6.21     Transmembrane    96 - 112 ( 88 - 116)
    INTEGRAL    Likelihood = -5.79     Transmembrane    249 - 265 ( 246 - 266)
    INTEGRAL    Likelihood = -2.87     Transmembrane    222 - 238 ( 222 - 238)
    INTEGRAL    Likelihood = -2.28     Transmembrane    279 - 295 ( 278 - 295)

15  ----- Final Results -----
        bacterial membrane --- Certainty=0.7496 (Affirmative) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

20  The protein has homology with the following sequences in the GENPEPT database:
    >GP:BABL3498 GB:AB026634 RNA polymerase [Flammulina velutipes]
    Identities = 83/336 (24%), Positives = 150/336 (43%), Gaps = 40/336 (11%)

    Query: 152 ILLLIAPVSIQKNR-VYNFVQNLYFEEVINNYFEENPVKIKEKSLIK-----FLMTIS 205
    IL L SI NR ++ N ++ N+P+ ++ K K L+I F++ +S
    Sbjct: 133 ILFLYLLYSILINRFILKWLDSGLIYKINWFWHNMKIKHINKMLVINIKFNPFIPIKLS 192

    Query: 206 FVFVIDPAMVRL-----LNFNIKFTSLICSAILLAWLYQN-----KSVIEPFL 249
    ++ I ++ L ++ N+P+I+ I I ++ S+ F
    Sbjct: 193 IITIIGISIMELFGIFGINDIRIIINVLKTINGSKHILTIINMDQVSVLENSHTIFY 252

    Query: 250 LKGLVYIFPIFIATLIGNLKN-ELGILETPLLFIISIFFTMDRIIALSKEMRDLI--ISKS 306
    + L+I+ IF L N+KN + I ++L+I IF I ++DE+ ++K
    Sbjct: 253 INLLIIFLIPISLLYRNVDINIDINIKRWIILYILFLINIIFFPHIYIKGLMGNLNY 312

    Query: 307 ILFFYDHENIKPSILLSEIKIYKLENVDIGE---LALVRQMVIRLRLELSSEFSLSDI 363
    IL Y D I S+ L ++K L+ ++I + V+ + I+ ++E L + I
    Sbjct: 313 ILDYMDLHIVNSLFLNKFQV-KLRINIYKSYSTVTYKDLKIKSKIEERSELDIKLI 371

    Query: 364 YMKNG-YKCIQFQGVNYPINLE--LDKIPFYTLKILLESIFD---HNNQKIFIPKL 416
    K Q YR YI ++ N+ ++ E L P Y N +E+ + + F+ K+
    Sbjct: 372 IAKYGSYENYINSIE-NINIVDSFELKYNPEYINDSKPIEFPLMELEPLFRDHTSEFVGI 430

    Query: 417 YEEYIYLISLGEVEKAKEIL--KEVSDYLTESL 449
    YE L + K+L KE+ DY+ +L
    Sbjct: 431 YENLNSTNEKLEFLANKDILSENKEIFDYVLQLNL 466
  
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 281

A DNA sequence (GBSx0308) was identified in *S.agalactiae* <SEQ ID 901> which encodes the amino acid sequence <SEQ ID 902>. Analysis of this protein sequence reveals the following:

```

Possible site: 37
5  >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
        bacterial cytoplasm --- Certainty=0.3307 (Affirmative) < succ>
  
```

-356-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 282

- A DNA sequence (GBSx0309) was identified in *S.galactiae* <SEQ ID 903> which encodes the amino acid sequence <SEQ ID 904>. This protein is predicted to be purK (purK). Analysis of this protein sequence reveals the following:

Possible site: 34
>>> Seems to have no N-terminal signal sequence
----- Final Results -----
bacterial cytoplasm --- Certainty=0.0334 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 20 A related GBS nucleic acid sequence <SEQ ID 9461> which encodes amino acid sequence <SEQ ID 9462> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

- >GP:CAA04376 GB:AJ000883 purK [Lactococcus lactis]
Identities = 208/347 (59%), Positives = 258/347 (73%), Gaps = 3/347 (0%)
25 Query: 14 NSFTIGIIGGGQLGQMAIAAIYMGHKVITLDPASDCPASRVSEVIVAPDDVEALGT 72
N+ +TIGIIGGGQLGQMAIAA YMGHKVITLDP +C A++VS E+IVAPDDVE L
Sbjct: 4 NTSQIIGGGQLGQMAIAAQMGHKVITLDPHFWCSAAKVSDELIVAPDDVENLLR 63
30 Query: 73 LAARCDVITYEFENVADGLDAVVSAGQLPQGTDLRLSQNRIFEKDFLANAGVTVAPY 132
LA CDV+TYEFENV A L + ++PQS LL I+QRR FEK+EL N+A V VAP+
Sbjct: 64 IAYACDVITYEFENVSAKALHEISGCVRIFGQIRLLBITQRRFEKEFLTNEAKVNAVW 123
35 Query: 133 KVVTSSLDLEGLDLTKTYVLKFTATGCGYDGHGQKVIRISADLPEAQQANSAQCVLREPVN 192
++V S+ L +T+ VLKT TCGYDGHGQ V+ + E L A+ L ++CVLE+P++
Sbjct: 124 QLVDSAEKLPET- VTRK+VLKTTTGGYDGHGQVLTDEKLSAKSLNELSECVLEDFIS 182
40 Query: 193 FDEISIVTSGNQQDVTVFPVQPAIHRANILSKTIVPARISDQLADAKEMAVQIAKGLQ 252
P+ EISVI+SGNQ + VFP+ EN HR NIL +TI PARIS ++ + A ++A IA+KL+
Sbjct: 183 FERIEISVILSGNGHETVVPFLANHEHRENILKQTIQPARISAEITENAVKIATISAEKLE 242
45 Query: 253 LSGTLCVENFATAD-DIIVNEIAPRPHNGHYSIEACDPSQDTHTLGLVGLAPLPKILH 311
LSG LCVEMF TAD I VNE+APRPHNGSH++IEACD+QND HI G+LG LP KL
Sbjct: 243 LSGVLCVENFMTADQIYVNEIAPRPHNGHIFTIACDQNFQDLHIKGIIGEDLPEPKLL 302
Query: 312 APANMFWVLGCHVQQAIDIVANPSAHLIMYGKLSAKGNRKGHVTV 358
P+M NVLGCHV+ ++ H H YGK +AKGNRKGHVT+
Sbjct: 303 KPAILMLNVLGCHVVEAKKLNHSHADWHQHDYCKADAKGNRKGHVTI 349

- 50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 905> which encodes the amino acid sequence <SEQ ID 906>. Analysis of this protein sequence reveals the following:

Possible site: 34
>>> Seems to have no N-terminal signal sequence

- 55 ----- Final Results -----

-357-

bacterial cytoplasm --- Certainty=0.0334 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below:

Identities = 344/369 (93%), Positives = 353/369 (95%)

```

Query: 1  MRNKEKSRQSQMNSFKTIGTIGGGGLGQMMAIAATYMGHKVITLDPASDCPASRVSEVI 60
      MRNKEKSRQSQ +NSFKTIGTIGGGGLGQMMAIAATYMGHKVITLDPASDC PASRVSEVI
10  Sbjct: 1  MRNKEKSRQSQVNSFKTIGTIGGGGLGQMMAIAATYMGHKVITLDPASDCPASRVSEVI 60

Query: 61  VAPYDDVEALGTLAARCDVLTFRFENVADGLDAVVSAGQLPQGTDLRLISQNRIFEKDF 120
      VAPYDDVEALG LAAACDVLTFRFENVADGLDAVVSAGQLPQGTDLRLISQNRIFEKDF
12  Sbjct: 61  VAPYDDVEALGTLAARCDVLTFRFENVADGLDAVVSAGQLPQGTDLRLISQNRIFEKDF 120

Query: 121 LANKAGVTVPAPYKVTSSLDLGLDLTKTYFLKATGQYDGHGQKVIASADLPEAQQLA 180
      LANKAGVTVPAPYKVTSSLDL GLDLTKTYFLKATGQYDGHGQK+IRASADLPEAQQLA
18  Sbjct: 121 LANKAGVTVPAPYKVTSSLDLGLDLTKTYFLKATGQYDGHGQKIIRASADLPEAQQLA 180

Query: 181 NSAQCVLEEFVNFLEISVIVSGNGQDVTVFFVQENIHRNNILSKTIVPARISDQLADKA 240
      NSAQCVLEEFVNFLEISVIVSGNG+DVTVFFVQENIHRNNILSKTIVPARISDQLADKA
20  Sbjct: 181 NSAQCVLEEFVNFLEISVIVSGNGQDVTVFFVQENIHRNNILSKTIVPARISDQLADKA 240

Query: 241 KENGAVQIAKKLQLSGTLGVEMFATADDIIVNEIAPRPHNSGRYSIEACDFSQFDTHILGV 300
      K+ AVQIAKKLQLSGTLGVEMF TADDIIVNEIAPRPHNSG YSIEACDFSQFDTHILGV
22  Sbjct: 241 KKTAVQIAKKLQLSGTLGVEMFATADDIIVNEIAPRPHNSGRYSIEACDFSQFDTHILGV 300

Query: 301 LGAPLPIIKLHAPAVMPFVLGCHVQQAIDHVAQNPSAHLHMYGKLEAGHNRKMGHVTVFS 360
      LGAPLP I+LHAPAVM NVLGGHVVQA D+VA+NPASHLHMYGKLEAGHNRKMGHVTVF+
30  Sbjct: 301 LGAPLPIQLHAPAVMLNVLGCHVQQAIDHVAQNPSAHLHMYGKLEAGHNRKMGHVTVFA 360

Query: 361 DVPSVEVEEF 369
      DEV+EF
32  Sbjct: 361 KDADVEVEEF 369

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 283

A DNA sequence (GBSx0310) was identified in *S. galactiae* <SEQ ID 907> which encodes the amino acid sequence <SEQ ID 908>. This protein is predicted to be phosphoribosylaminoimidazole carboxylase catalytic subunit (purE). Analysis of this protein sequence reveals the following:

Possible site: 45
 >> Seems to have no N-terminal signal sequence

```

45  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3572 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

50 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAE12462 GB:Z99107 phosphoribosylaminoimidazole carboxylase I
 [Bacillus subtilis]
 Identities = 106/162 (65%), Positives = 128/162 (78%)

```

55  Query: 33  MQPIISIKSGESDWTMTQKTAVELDNFGLAYEKKVVSARHTEDLMFHKHAEARGGIGKI 92
      MQP++ IINGS SSW TW+ ++LD + YEKVVSARHTEDMF++AE AR RGK+
      Sbjct: 1  MQPLVGIINGSSTSWETMKHACDILDLNVPYEKKVVSARHTEDMFYSAETARERGIKV 60

Query: 93  ILAGGAGGAHLPMQVAAKTTLPVIGVFPVKSRLSGLDLSLIVMQPGVPVATMAIGAG 152
      ILAGGAGGAHLPGM AAKTTLPVIGVFPV+S+AL+G+DSL SIQMGPGVPVAT +IG+AG
60

```

-358-

Sbjct: 61 I IAGAGGAHLPGMTAKTTLPVIGVPVQS KALNMDSLISIVQMPGGVPVATTS IGKAG 120

Query: 153 ATNALATRLILSTEDQNLADALAHFHESQKIAEBSNELI 194

A N A L A + I L S D + I A L E + R S S + + L +

Sbjct: 121 AVNAGLLAAQILSAPFDLARKLDERRENTYKTIVLESSTQLV 162

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 909> which encodes the amino acid sequence <SEQ ID 910>. Analysis of this protein sequence reveals the following:

Possible site: 57
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -3.08 Transmembrane 36 - 52 (34 - 52)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2232 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAA04375 GB:A7000883 purE [Lactococcus lactis]
 Identities = 105/158 (66%), Positives = 131/158 (82%)
 Query: 46 ISIINGSKSIWATMKTAEVLNFGIAYEKKVSAHRTFPLMFHAEBAERGRGKIKI IAG 105
 ++IING SIWATM++TA++LD+FG+AYEKKVSAHRTF LM + + +AR RG K+I IAG
 Sbjct: 4 VAIINGCSSIWATMKTAKILDDFGLAYEKKVSAHRTFALMAEFSSQARERGVK I IAG 63
 Query: 106 AGGAHHLPGMVAAKTTLPVIGVPVKSRLSGDLSYISIVQMPGGVPVATMAIGAGATNA 165
 AGGAHHLPGMV+A+T +PVIGVP+KSRLSGDLSYISIVQMP GVPVATMAIGAGATNA
 Sbjct: 64 AGGAHHLPGMVAQTLVPVIGVPIKSRLSGDLSYISIVQMPAGVPVATMAIGAGATNA 123
 Query: 166 ALATRLILSTEDQNLADALAHFHESQKIAEBSNELI 203
 AL AL++L+ ++NL L + ++ EBS+ L+
 Sbjct: 124 ALFALQLLANTNENLIXQLLVYRAAQMVEESNKALL 161

An alignment of the GAS and GBS proteins is shown below:

Identities = 162/169 (95%), Positives = 164/169 (96%), Gaps = 1/169 (0%)
 Query: 27 PLYINIMQ-PIISIIIMSGSKSDWTTMKTAEVLNFGIAYEKKVSAHRTFPLMFHAEBA 85
 PL + IM+ PIISIIIMSGSKSDW TMKTAEVLNFGIAYEKKVSAHRTFPLMFHAEBA
 Sbjct: 35 PLCLINKPTIISIIIMSGSKSDWATMKTAEVLNFGIAYEKKVSAHRTFPLMFHAEBA 94
 Query: 86 RRGKIKII IAGAGGAHLPGMVAAKTTLPVIGVPVKSRLSGDLSYISIVQMPGGVPVAT 145
 RRGKIKII IAGAGGAHLPGMVAAKTTLPVIGVPVKSRLSGDLSYISIVQMPGGVPVAT
 Sbjct: 95 RRGKIKII IAGAGGAHLPGMVAAKTTLPVIGVPVKSRLSGDLSYISIVQMPGGVPVAT 154
 Query: 146 MAIGAGATNAALATRLILSTEDQNLADALAHFHESQKIAEBSNELI 194
 MAIGAGATNAALATRLILSTEDQNLADALAHFHESQKIAEBS EELI
 Sbjct: 155 MAIGAGATNAALATRLILSTEDQNLADALAHFHESQKIAEBS SKILI 203

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 284

A DNA sequence (GBS0311) was identified in *S.galactiae* <SEQ ID 911> which encodes the amino acid sequence <SEQ ID 912>. This protein is predicted to be phosphoribosylglycinamide synthetase (purD). Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----

-359-

bacterial cytoplasm --- Certainty=0.1966 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database:

>GP:CA04374 GB:AJ000883 purD [Lactococcus lactis]
 Identities = 236/419 (56%), Positives = 298/419 (70%), Gaps = 7/419 (1%)

10 Query: 1 MKLLVGGSGREHAIAKKLLASKDVQVVPAPGNDGNTLDGLDLVNIIGISEHSRLIDFVK 60
 MK+LV+GSGGREHA+AKK + S V++VPVAPGN GM DG+ +V+I + +L+ F +
 Sbjet: 1 MKILVIGSGGREHALAKKFMESQVVEVPVAPGNSGMEKDGQIQVHISELSNDKLVKFAQ 60

Query: 61 ENEIAMTWLIGPDDALAAGIVDGFPSAGLRAPGPTKAAALEWSKDFAKEIMVKYNVPTAA 120
 I T +GP+ AL G+VD F A L RGP K AALE SKDFAK IM KY VPFA
 15 Sbjet: 61 NQNIQLTFVGPETALMAGVVDAPFAIKELPIPGFNMAAALBESKDFAKSIMKYGVPTAD 120

Query: 121 YGTFSDPEKAKAYIEEQCAPIVVKADGLALGKGVVAETVEQAVEAAGNLLDNKFGDSG 180
 Y TF E A AY++E+G P+V+KADGLA GKGVA +E A A ++ F S
 20 Sbjet: 121 YATFDSLEPALAYLDEKGVPLVIKADGLAAGKGVTVAFDIETAKSALADI-----FSGSQ 175

Query: 181 ARVVEEFLDGEFSLFAFANGDKFYIMPTAQCHKRAYDQDKGLATGGMGAYAPVPHLPQ 240
 +VVIEEFLDGEFSLF+ F + K Y MP AQDHKRA+D DKG MTGGMGAY+PV H+ +
 Sbjet: 176 GKVVEEFLDGEFSLFSDIHGDGILYPMPIAQDHKRAFDDGKPNYTGGMGAYSPVLIHSK 235

25 Query: 241 SVVDTRAVETIVKPVLEGMIAEGRPVLGVLYAGLILMADGPKVIEFNSRFGDPETQIILR 300
 VV+ A+E +VKP + GMI EG+ + GVLVAGLILT DG K IEFN+RFGDPETQ+LFR
 Sbjet: 236 BVNLSALEKVVKPTVAGMIEBESKPTGVLYAGLILTEDGKTIIEFNARFGDPETQVVLFR 295

30 Query: 301 LTSDPAQNIDIMMIEEFYITWQKDGVTGLGVVVASGEPYLDYEGKGVPLPEKIDGDIITY 360
 L SD AQ I DI+ G EF + W + GVLGVVVA+EGYP + G+ LFE +G + YY
 Sbjet: 296 LMSDLAQALIDILAGNEFTLEWLESGVTGLGVVVASGEPYSAKGLLILPEFPG-LNVYY 354

Query: 361 AGAKFAENSKALLSNKGRVIMLVTTEDSVKAGQDITVLAQQDITGLFRNDIGSKAI 419
 AG EN+ L+S+GGRVY++ T +VK+ Q + +L+ + + G FR+DIGS+AI
 35 Sbjet: 355 AGVSNENNQ-LISSGGRVYLSETGEDVKSTKLLYKLLDLENLGGFFFRHDIGSKAI 412

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 913> which encodes the amino acid sequence <SEQ ID 914>. Analysis of this protein sequence reveals the following:

40 Possible site: 35
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.80 Transmembrane 5 - 21 (5 - 21)
 ----- Final Results -----

45 bacterial membrane --- Certainty=0.1319 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

50 >GP:CA04374 GB:AJ000883 purD [Lactococcus lactis]
 Identities = 236/419 (56%), Positives = 301/419 (71%), Gaps = 7/419 (1%)

Query: 50 LKLLVGGSGREHAIAKKLLASKDVQVVPAPGNDGNTLDGLDLVNIIVSEHSRLIAFAK 109
 +K+LV+GSGGREHA+AKK + S V++VPVAPGN GM DG+ +V+I + +L+ FR+
 55 Sbjet: 1 MKILVIGSGGREHALAKKFMESQVVEVPVAPGNSGMEKDGQIQVHISELSNDKLVKFAQ 60

Query: 110 ENEIAMTWLIGPDDALAAGIVDGFPSAGLRAPGPTKAAALEWSKDFAKEIMVKYNVPTAA 169
 I F+GP+ AL G+VD F A L RGP K AALE SKDFAK IM KY VPFA
 Sbjet: 61 NQNIQLTFVGPETALMAGVVDAPFAIKELPIPGFNMAAALBESKDFAKSIMKYGVPTAD 120

60 Query: 170 YGTFSDPEKAKAYIEEQCAPIVVKADGLALGKGVVAETVEQAVEAAGNLLDNKFGDSG 229
 Y TF E A AY++E+G P+V+KADGLA GKGVA +E A A ++ F S
 Sbjet: 121 YATFDSLEPALAYLDEKGVPLVIKADGLAAGKGVTVAFDIETAKSALADI-----FSGSQ 175

Query: 230 ARVVEEFLDGEFSLFAFANGDKFYIMPTAQDHKRAFDDGKPNYTGGMGAYAPVPHLPQ 289

-360-

```

+VVIEEFLDGESEFLS+F + K Y MP AQDHKRAFD DKGENTGGMAY+PV H+ +
5  Sbjct: 176 GKVIVIEEFLDGESEFLSFDHDKGLIYMPFIADQHKRAFDDEKGENITGGMAYSPVLHISK 235
Query: 290 SVVDITAVEMIVRPVLEGWVAERPFYLGVLVGLILITADGPKVIEFNSRFGDPETQIILPR 349
VV+ A+E +V+P + GM+ EG+ + GVLY GLIIT DG K IEFN+RFGDPETQ+LPR
10 Sbjct: 236 EVNVEALEKVVVKTAVGMIEKSKSFTGVLYAGLILITEDGKVTIEFNARFGDPETQVVLPR 295
Query: 350 LITSDFAQNIDDIMMGIEFPYITWQKDGVTGLGVVVASSEGYF DYKGVPLPEKTDGDIITYY 409
L SD IQ I DI+ G EP + W + GVTGLGVVA+EGYP + G+ LPE +G + YY
10 Sbjct: 296 LKSDLAQAIIDILAGNEPITLWLESQVTLGVVVVAERGPSQAKLGLILPEIPEG-LANVY 354
Query: 410 AGVKFSESEILLNSGRVYMLVTTEDSVKAGQDKIYTLAQDQTTGLFYRNDIGSKAI 468
AGV +EK++ L+S+GRVY++ T + VK+ Q +Y +L + + G FYR+DIGS+I+
15 Sbjct: 355 AGVSKNENQ-LISSGRVYLVSSETGEDVKSTKILYRKLDKLENDGFFYRHDIGSRAI 412

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 399/421 (94%), Positives = 408/421 (96%)
20 Query: 1 MKLLVVGSGGRBHAIAKLLASKVDQVFVAPGNDGMLTGLDLVNIIGISEHSRLIDFVK 60
+KLLVVGSGGRBHAIAKLLASK VDQVFVAPGNDGMLTGLDLVNI +SHSRLI F K
Sbjct: 50 LKLLVVGSGGRBHAIAKLLASKGVDQVFVAPGNDGMLTGLDLVNIIVVSEHSRLIAFAK 109
Query: 61 ENIISAWLIGPDDALAGIVDGFNSAGLRAGPTKAAAELEWSKDFAKIEMVKNVPTAA 120
ENI+M IGPDALAGIVD FNSAGLRAGPTKAAAELEWSKDFAKIEMVKNVPTAA
25 Sbjct: 110 ENIISWAFIGPDDALAGIVDGFNSAGLRAGPTKAAAELEWSKDFAKIEMVKNVPTAA 169
Query: 121 YGTFSDFEKAKAYIEQGAPIVVKADGLALGKGVVVAETVBOAVEAQAQMLLNKFGDGG 180
YGTFSDFEKAKAYIEQGAPIVVKADGLALGKGVVVAETVBOAVEAQAQMLLNKFGDGG
30 Sbjct: 170 YGTFSDFEKAKAYIEQGAPIVVKADGLALGKGVVVAETVBOAVEAQAQMLLNKFGDGG 229
Query: 181 ARVVEEFLDGESEFLSFAFANGDKFYIMPTAQDHKRAYDGDGKLNTGGMAYAPVPHLQ 240
ARVVEEFLDGESEFLSFAFANGDKFYIMPTAQDHKRA+DGDGK NTGGMAYAPVPHLQ
Sbjct: 230 ARVVEEFLDGESEFLSFAFANGDKFYIMPTAQDHKRAFDGDKGNTGGMAYAPVPHLQ 289
35 Query: 241 SVVDITAVETIVKPVLEGMIARERFYLGVLVGLILITADGPKVIEFNSRFGDPETQIILPR 300
SVVDITAVE IV+PVLEGW+AEGRPFYLGVLV GLIITADGPKVIEFNSRFGDPETQIILPR
Sbjct: 290 SVVDITAVEMIVRPVLEGWVAERPFYLGVLVGLILITADGPKVIEFNSRFGDPETQIILPR 349
Query: 301 LITSDFAQNIDDIMMGIEFPYITWQKDGVTGLGVVVASSEGYF DYKGVPLPEKTDGDIITYY 360
LITSDFAQNIDDIMMGIEFPYITWQKDGVTGLGVVVASSEGYF DYKGVPLPEKTDGDIITYY
40 Sbjct: 350 LITSDFAQNIDDIMMGIEFPYITWQKDGVTGLGVVVASSEGYF DYKGVPLPEKTDGDIITYY 409
Query: 361 AGAKFASNSKALLNSGRVYMLVTTEDSVKAGQDKIYTLAQDQTTGLFYRNDIGSKAIKE 421
AG KF+ENS+ LLNSGRVYMLVTTEDSVKAGQDKIYTLAQDQTTGLFYRNDIGSKAI+E
45 Sbjct: 410 AGVKFSESEILLNSGRVYMLVTTEDSVKAGQDKIYTLAQDQTTGLFYRNDIGSKAIR 470

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 285

50 A DNA sequence (GBSx0312) was identified in *S. galactiae* <SEQ ID 915> which encodes the amino acid sequence <SEQ ID 916>. Analysis of this protein sequence reveals the following:

```

Possible site: 36
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -1.28 Transmembrane 235 - 251 ( 235 - 251)
55 ----- Final Results -----
bacterial membrane --- Certainty=0.1510 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
60

```

The protein has homology with the following sequences in the GENPEPT database:

-361-

>GP:AA23257 GB:M81878 unknown [Clostridium perfringens]
 Identities = 66/258 (25%), Positives = 119/258 (45%), Gaps = 9/258 (3%)

5 Query: 1 MTIYDQIESALDINTDLEREIACYPMQPISKDALASTIVTQQLHISQAALTRFAKKCOF 60
 M I +Q+E+ T E+ + Y + + +I+ K+ + +A +TRF KK GP
 Sbjct: 1 MGILEQLENPKFKATKSEKTLIEYTKSDLNITIKSISITAKSGVGEATITRFTKILG 60

10 Query: 61 KGYREFVFEYLES-HETISQQLGLQNDNTKKVFMNYQEMISKADI-----IDEBQL 112
 G+++F K + + L + V +N+ S +I ID + +
 Sbjct: 61 NGQDFKVTAKIEISNKKNTISILNVEHDESVTETANKMLASSINILQTVKQIDILDM 120

15 Query: 113 LEVSHMIEQADRVFYFYGGSSSLVAKEFKIRLMRLGVICELDDTDSFWSNINSVNDRL 172
 + +I A RVYF G G S + A + +MR+G + D+ + +SI ND +
 Sbjct: 121 CKCRDLIMNAKRVYFYGIGYGLAATDINMKFMRIGFTTVPVVTDSDHVMVMSITNDIDV 180

20 Query: 173 VIAFSLSGNTNSVIGALKIASCHGAKTVLFTK-QPHTIDYAFDKIIQVASARHLYGNRI 231
 ++A S SG T VI +K A +G K + T+ + + D + SA + I
 Sbjct: 181 IVAISGSGTTRKVKITVKQAKENGFKIITHFSDSDNPLRLKSDYELTYTSAETITGSI 240

25 Query: 232 SPQIFMLIMVDLIVAQFL 249
 S +IP + ++D++Y + +
 Sbjct: 241 SSKIFQIFLLDLLLYTEVI 258

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 917> which encodes the amino acid sequence <SEQ ID 918>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -4.88 Transmembrane 243 - 259 (242 - 261)

30 ----- Final Results -----
 bacterial membrane --- Certainty=0.2954 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related sequence was also identified <SEQ ID 9093> which encodes the amino acid sequence <SEQ ID 9094>. Analysis of this protein sequence reveals the following:

Possible cleavage site: 56
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -4.88 Transmembrane 239 - 255 (238 - 257)

40 ----- Final Results -----
 bacterial membrane --- Certainty= 0.295 (Affirmative) < succ>
 bacterial outside --- Certainty= 0.000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty= 0.000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 138/263 (52%), Positives = 199/263 (71%), Gaps = 2/263 (0%)

50 Query: 6 QIESALDINTDLEREIACYPMQPISKDALASTIVTQQLHISQAALTRFAKKCOFGKYRE 65
 +IE++L+ MT LE+ IA +F+ ++ L ++ + K+LHISQAALTRFAKKCOF GYR
 Sbjct: 14 KIEASLEHMTSLERGLDAFFITTDLPQLTASEIVKRLHISQAALTRFAKKCOFTGYRA 73

Query: 66 FVFEYLSHETISQQLGLQNDNTKKVFMNYQEMISKADIIDEBQLLEVSHMIEQADRV 125
 F F+YL S + + + TK+V M+Y +I+K+ +++ER+L ++ +I+ ++RV
 55 Sbjct: 74 FAFDYLSLQESQETPQSHLGLTKRVLMYDALINKTYELVNEKILNLAKLIDSSERV 133

Query: 126 FYFYGGSSSLVAKEFKIRLMRLGVICELDDTDSFWSNINSVNDRLVIAFSLSGNTNSV 185
 YF+GKGSS LVA+R K+R MRIG+IC+A DTD F+W NS+VN+ C+V FSLSG TNSV
 Sbjct: 134 FYFGKGSSGLVAREMKLFPNRLGLCDAYSDTGTWANSLVNENCINPFGSLSGNTNSV 193

60 Query: 186 IGALKIASCHGAKTVLFTKQPHT-IDYAPDKIIQVASARHLYGNRIISQIFMLIMVDII 244
 I AL AS GAKTVL T T D + D II V+S L YGNR+SPQ P+LIM+DII
 Sbjct: 194 ITALHQAORGAKTVLLTIDNQTDFDSDLD-IIPVSTHQLHYGNRVSQFPFLIMDII 252

Query: 245 YACFLDINKIKKERIPRETIQIR 267
 YA L I+K K3+IF+ TII +
 Sbjct: 253 YAYVLADKKHKKIKKIPRETIQIR 275

SEQ ID 916 (GBS320) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 40 (lane 5; MW 33kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 85 (lane 7; MW 58kDa) and in Figure 160 (lane 7 & 8; MW 58kDa).

GBS320-GST was purified as shown in Figure 224, lane 3-4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 286

A DNA sequence (GBS0313) was identified in *S. agalactiae* <SEQ ID 919> which encodes the amino acid sequence <SEQ ID 920>. This protein is predicted to be xylan esterase 1 (cephalosporin-C). Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4981(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB68821 GB:AF001926 xylan esterase 1 [Thermoanaerobacterium sp.
 'JW/SL YS485']
 Identities = 133/299 (44%), Positives = 188/299 (62%), Gaps = 1/299 (0%)

Query: 5 NSLDDMREYLGQDQIPEDFDDPMKQTMKYQG-NIEYRLDKKDHITFACAYDLHFKGSN 63
 N L +REY G + PEDFD+H + + + + L + F ++E+ YDL+F G
 Sbjct: 6 NPLQLREYLTGTNCPEDFDEYRNALDEMRSDPKTELKESSPQVSPACDLYFTGVR 65

Query: 64 NSITVAKCLFPKTKNKPVPVFFPHGYQNCSPDMSQIAYVAAGYGVVSDVROQAGQGSQ 123
 + +AK + PKT +D + FHGY + S DW-D+LNTVAAG+ VV+NDVROQ GSQD
 Sbjct: 66 GARIHAKYIKPTBGKHPALIRFHGYSSNSCDHNDKINYVAAGFTVAVMDVROQAGQGSQ 125

Query: 124 KGHFDGHTVKGQIVRCMISGPNHLPYKDIYLDVPLQIDLIATLESVDNQLYSYGVSGQG 183
 G G T+ G I+RG+ + + + + I+LD QL I+ + VD + + + G SQGG
 Sbjct: 126 VGVVONTNCHIIIRCLDDDDNMLPHIIFLDVTAQLAGIVQMEVDEDRVGVSGPQGG 185

Query: 184 ALALAAALNPKIVKIVAVVPLSDFRVRLDGGVSEYDELPRYFKYSDPPHKTENVL 243
 L+L RA5 P+ K V+ YPLSD+RV DL Y E+ YF+ DP H+ EN V
 Sbjct: 186 GLSLACALAEPRVRKVVSEYPLSDYKGVWDLDAKNAQYETIDYPRFLDPRHRENEVF 245

Query: 244 KTLAVIDVNFPAHRIISCPVVLITALKADDICPPSTQFAIFNRLTSTKGHILLDPYGHDFM 302
 L YIDVN A RI V+ + L D +CPPST FA +N + E K + PDYGH+FM
 Sbjct: 246 TKLGIVIDVNLAKRIKGDVLMCVGLMDQCPPSTVFAAYNLIKQKKDKVYPDYGHDFM 304

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 287

A DNA sequence (GBSx0314) was identified in *S. agalactiae* <SEQ ID 921> which encodes the amino acid sequence <SEQ ID 922>. Analysis of this protein sequence reveals the following:

```

Possible site: 35
>>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -5.73    Transmembrane    128 - 144 ( 126 - 145)

----- Final Results -----
      bacterial membrane --- Certainty=0.3293 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAA23256 GB:M81878 unknown [Clostridium perfringens]
Identities = 78/160 (48%), Positives = 110/160 (68%)

Query: 131 CLTIGTIGGGCLIIDKTVFHGFSNSACEVGYMHLSDGDFQDLASTTALIDAVKAGHDEI 190
          CLTIGTIGGG LIID V HGPSNSA E+GYM ++ + QD+AS +AL+ +VA G E
Sbjct: 18 CLTIGTIGGALLIDGKVLHGFNSAGEIGYMNNGENIQDASASALKNVLRKGVEP 77

Query: 191 SRWDGRRIFQEAQKNGEKCIASIDRMNVLGGQIANMVTYVNPKEKVLGGGIDAKQDIYQ 250
          S DGR + + G+ C +++ + L G1+N+VY++NPE VVLGGGIMA+++ +
Sbjct: 78 SSIDGRYVLNDYENGDLCKEVEKRLADNLALGISNIVYILNPEVVLGGGIMARESEVR 137

Query: 251 DKLSGLKRNVLVTSLARKTAIVFAQHNCAGMLGAYYHFK 290
          + SL++ L+ S+ T I FA+ +N AGM GAYY+FK
Sbjct: 138 FLIENSIRKYLIESVYNNTKIAFAKLNTAG+KGAYYHFK 177

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 923> which encodes the amino acid sequence <SEQ ID 924>. Analysis of this protein sequence reveals the following:

```

Possible site: 22

>>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -4.30    Transmembrane    128 - 144 ( 127 - 145)
      INTEGRAL    Likelihood = -0.11    Transmembrane    227 - 243 ( 227 - 243)

----- Final Results -----
      bacterial membrane --- Certainty=0.2720 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:BAB04516 GB:AP001509 glucose kinase [Bacillus halodurans]
Identities = 97/291 (33%), Positives = 155/291 (52%), Gaps = 14/291 (4%)

Query: 5 LAIDIGGTAKYGLISRTGDLLEKEBMATEAYKGPSPILEKVKGLVKTQDQMDLGAVAL 64
          + ID+GKT IK L+S+ G+++ +E TRA +G ++ K+ L+ D AG+ I
Sbjct: 3 VGIDLGDTKIKAAVLSDAGEIISVQECPTMAQGEVNNRMSLTEKTVTHDHPAGIGI 62

Query: 65 SSAGMNVNPDGEIFYAGQIPINVAGTQPKKEISETTGLFCEVENDVNCALEAISGSAS 124
          + G ++ EG I + P + P + +E F P +++ND N A LAEA+ GS +
Sbjct: 63 GAPGFLSSTGTIL-SPNPLGWDHILHLVDRFQBPQCPVKLNDANVAALAEALLGSQ 121

Query: 125 DYPVALCLTIGTIGGCLFNFSQVHSGNSHSAEFG-----YHLSDGQFODLAS 174
          + LTI TGIGS + + + HG+S A E+G + +L+ G + L+S
Sbjct: 122 GFTSVFYLTISTGIGGGYVLDGSIHVHGSADYAGIEGNMIVQPNGYQHNINPGSLEGLAS 181

Query: 175 TTLVQEVFLAYGGDIDISQMDGRRIFQEAQAGDAICAAISKQVDYLAGQIANICITYVNN 234
          TA+ + + +G + R +F+Q + GD + + +DYL GIANI + +H+
Sbjct: 162 GTAIGRMARERFG---VGGSTREVFQDILRGDHQMRVLBEAMDYLAIGIANIANTINPD 238

```

-364-

Query: 235 VVVLGGGIMAQKDYLAADKLKLTALDLYSLVSSLAKTQKLFASHGNNNGILGA 285
 V VLGGGM D + +K + YL IA+ T + A G +G+GA
 Sbjct: 239 VVVLGGGVMADDLILPIVKEKVSRYLPGLAQSTTVIKAKLGDSGVILGA 289

5 An alignment of the GAS and GBS proteins is shown below:

Identities = 192/292 (65%), Positives = 237/292 (80%)

Query: 1 MTKTVADIGGTMIKHGIVDNLGCIIVEASELATEAYKGGGILQKVCQI IDWYLABSSID 60
 M +AIDIGGT IK+G++ G ++R E+ATEATKGGP IL+KV ++ Y + +
 10 Sbjct: 1 MKETLAIDIGGTALTYGLISSETGDLLEKEEMATEATKGGPSILSKVGLVKTYQDQIDLA 60

Query: 61 GIAISGAGMVDPEBCIFYSGPQIPNAGTQFKKVLDTYQVKTETENDVNCAGLAAVS 120
 G+AISGAGM+PDEG IFY+GPQIPNAGTQFKK +E+T+ + E+ENDVNCAGLAA+S
 15 Sbjct: 61 GVAISGAGMVDPEBGRIFPAGPQIPNAGTQFKKEIEETPGLFCEVENDVNCAGLAAIS 120

Query: 121 GSAKDSIALCLTIGTGIGGCLLIIDKTVPHGFSNAGCEVGMHLSGDFQDLASTTALA 180
 GSAKD +ALCLTIGTGIGGCL+ + VPHS S+SACEVSY+HLSDG FQDLASTTALA
 15 Sbjct: 121 GSAKYFPALCLTIGTGIGGCLLPHS VPHGSHSAGCEVGYLHLSDFQDLASTTALVQ 180

Query: 181 DVAKAGDEBSINWDRRIFQBAKKGNEKCIASIDRMINTYLGQGIANNMVVYNPHKVVVLG 240
 +V R+GD+IS+WDGRIF++AK G+ CIA+I + ++YLQGGIAN+ VVNP VVLGG
 20 Sbjct: 181 EVVLAYGDDISWDGRRIFFQAKAGDAICIAAISQVDYLGQGIANICIVVFNVVVLG 240

Query: 241 GIMAQKDYLDQVLSGLKRLVLTSLAETKIVPAQHENGAGMLGAYYHFKR 292
 GIMAQKDYLDKX +L DV+SLA+RT + FR H N AG+LQAYVHFK +
 25 Sbjct: 241 GIMAQKDYLDKLTALDLYSLVSSLAKTQKLFASHGNNNGILGAYYHFKR 292

SEQ ID 922 (GBS331) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 60 (lane 2; MW 35.9kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 67 (lane 3; MW 61kDa).

The GBS331-GST fusion product was purified (Figure 209, lane 3) and used to immunise mice. The resulting antiserum was used for FACS (Figure 309), which confirmed that the protein is immunoreactive on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 288

A DNA sequence (GBSx0315) was identified in *S. agalactiae* <SEQ ID 925> which encodes the amino acid sequence <SEQ ID 926>. This protein is predicted to be an acylneuraminatase (nanA). Analysis of this protein sequence reveals the following:

40 Possible site: 18
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.0894 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CBA69950 GB:Y08695 putative acylneuraminatase [Clostridium
 certium]

10 Identities = 162/225 (72%), Positives = 191/225 (84%)

Query: 1 MKDLQKYGIIIPAFYACYDDGDCIPERVKALINYPIDKGVQGLVYNSGSGBCIYQSVAD 60
 M++L+KY+GIIIPAFYACYDD+G I PER + T Y IDKGV+GLYV GSGSGBCIYQS +
 55 Sbjct: 1 MNLEKYGKIIIPAFYACYDDGSGISPRTQMPTQYLLDKGVKGLVYVNSGSGBCIYQKEE 60

-365-

Query: 61 RKLVLNVMVARGKLTVAHVACNNTKDSVELAMHAEAGVDAIAAIPPIYFRLPEYAI 120
 RK+ LENVM VARGK+T+IAHV CNNT+DS ELA HAE+IGVDALIA+IPPIYF LP+Y+I
 Sbjet: 61 RKITLENVMKAVARGKITIAHVGCNNTRDSERLAHAEISIGVDAIASIPPIYFRLPEYAI 120

Query: 121 ADYWNITISQAAPQDTFIIYINIPQLAGVALTSDLYRKLQNPQVIGVKNSSMPVQDIQNFV 180
 A+YWN IS AAP TDFIIYINIPQLAGV L +LY+ML+NP+VIGVKNSSMPVQDIQ F
 Sbjet: 121 AEYWNDISQAAPNTDFIIYINIPQLAGVGLINLYKQLMKNPKVIGVKNSSMPVQDIQMFK 180

Query: 181 AIGGESHIVFNGPDEQFLGGRLMGAAAGIGGTGYMPELVLTLNQ 225
 I G+ +VFNGPDEQF+ GR+MGA GIGGTGY MPEL+L ++
 Sbjet: 181 DISGDESIVFNGPDEQFVAGRIMGADGGIGGTGYMPELFLAADK 225

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 927> which encodes the amino acid sequence <SEQ ID 928>. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0981(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 238/304 (78%), Positives = 263/304 (86%)

Query: 1 MKDLQKYGSIIPAFYACYDDKGDI CPERVWALINYPIDKGVQGLYVNGSSGECIYQSVAD 60
 M DL KYGSIIPAFYACYDD+G+I PERV+AIT Y+IDKGVQGLY+NGSSGECIYQSV D
 Sbjet: 1 MTLTKYQGIIPAFYACYDDQGNLS PERVWALITQYIDKGVQGLYINGSSGECIYQSVFD 60

Query: 61 RKLVLNVMVARGKLTVAHVACNNTKDSVELAMHAEAGVDAIAAIPPIYFRLPEYAI 120
 R+LVLNVM+VARGKLT+I FVACHNNTKDS+ELA H+E +GVDAIAAIPPIYFRLPEYAI+
 Sbjet: 61 RQVLNVMVARGKLTITN FVACHNNTKDSIELAHSERLGVDAIAAIPPIYFRLPEYAV 120

Query: 121 ADYWNITISQAAPQDTFIIYINIPQLAGVALTSDLYRKLQNPQVIGVKNSSMPVQDIQNFV 180
 ADYWN IS AAP TDFIIYINIPQLAGVALT LY+ ML N +VIGVKNSSMPVQDIQ F
 Sbjet: 121 ADYNAISQAAPNTDFIIYINIPQLAGVALTFSLYKTMLANKRVIGVKNSSMPVQDIQTFC 180

Query: 181 AIGGESHIVFNGPDEQFLGGRLMGAAAGIGGTGYMPELVLTLNQ 240
 AIGG+HIVFNGPDEQFLGGRLMGAAAGIGGTGY MPEL+L LKQLI DKLEKKA+ LQ+
 Sbjet: 181 AIGGEDHIVFNGPDEQFLGGRLMGAAAGIGGTGYMPELFLRLNQLIADKLEKKAALQY 240

Query: 241 TINDIITLCSGHNMYAVIKAVLEINEQLTIGSVRLPLASVTEDEKPIIKEAAEMIRHA 300
 TIN+II L S HGNMY VIK VL INE L IGSVR PLA + EED+ I +AA +I A
 Sbjet: 241 TINEITGLVSAHGNMYVKEIVLEINEQLDIGSVRSPLAEIIEEDVRVICQRAALINQA 300

Query: 301 KQPF 304
 K+ F
 Sbjet: 301 KTFP 304

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 289

A DNA sequence (GBSx0317) was identified in *S.galactiae* <SEQ ID 929> which encodes the amino acid sequence <SEQ ID 930>. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -9.45	Transmembrane	82 - 98 (79 - 111)
INTEGRAL	Likelihood = -6.85	Transmembrane	24 - 40 (21 - 52)
INTEGRAL	Likelihood = -5.26	Transmembrane	180 - 196 (172 - 200)

-366-

INTEGRAL Likelihood = -5.10 Transmembrane 160 - 176 (158 - 179)
 INTEGRAL Likelihood = -4.35 Transmembrane 110 - 126 (106 - 130)

----- Final Results -----

bacterial membrane --- Certainty=0.4779 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BA05827 GB:AP001514 unknown conserved protein in B. subtilis

[Bacillus halodurans]

Identities = 40/148 (27%), Positives = 74/148 (49%), Gaps = 4/148 (2%)

Query: 14 VDNFFMQGCVVFDLALNLLFMI-TCLPLVTIG--AAKISLYRTIWMQLEGD-QINILI 69

+++ F Q C+ ++ LA +NLL++ T L LV +G A +++ L + G+ +

Sbjct: 6 MSRFYQTCDWINKLAYINLNLMSGLTALGLVLGFLPATTAMPTVLREKTFNPFNDVAITR 65

Query: 70 LYIKHLKKEWPGQMLGLVLSILVVIIFDUTLLEYQIGFVSLFKITCYAPLLLTMTS 129

+ + K E+ + L L G V L L ++ E+ L G+ + L + YAPL+L ++T

Sbjct: 66 TFFQAYKNEFLKINLGAVALLYGAYILYFNMYTLGTVEGTVMVLGSLWYAPLILYITL 125

Query: 130 IYLFPMARYESMLLDTYKKSFMACLN 157

Y+ P Y + L +K + I+ +N

Sbjct: 126 FYIIPAYVHYNLKFQYIKTALIIGFVN 153

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 932> which encodes the amino acid sequence <SEQ ID 932>. Analysis of this protein sequence reveals the following:

Possible site: 24

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -14.86 Transmembrane 117 - 133 (108 - 139)

INTEGRAL Likelihood = -7.48 Transmembrane 30 - 46 (21 - 54)

INTEGRAL Likelihood = -6.90 Transmembrane 88 - 104 (83 - 105)

INTEGRAL Likelihood = -6.26 Transmembrane 165 - 181 (151 - 187)

INTEGRAL Likelihood = -5.89 Transmembrane 189 - 205 (182 - 207)

----- Final Results -----

bacterial membrane --- Certainty=0.6944 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BA05582 GB:AP001513 unknown conserved protein in bacilli

[Bacillus halodurans]

Identities = 59/194 (30%), Positives = 93/194 (47%), Gaps = 11/194 (5%)

Query: 17 SKWRASAAFLDILVFNLLFVL-SCLPLITIGV--AKMALYASLLDWHQGS-QLVITY 72

+K M+ + L+ NLL++L S + + +G V A +L+A W + + L TY

Sbjct: 8 TKIMKLFEMIRLVYLNLLMLLPSFISGIIILGVMPTASLPAVFRKWKQEDDFLPQTY 67

Query: 73 SSHFKYTFKSLRGLIELGIMTICLDLFLIRNQSGLVFPQFKVLGVNPLVPLVLP 132

+ FK FK +GL + I I Ld+ L+ S + Q + R+ R+ ++ LY

Sbjct: 68 LNFKRFKFKIANIAGTLVLVIGGILYLDVLLILGTSHWIGQLLMLGVNLSFYLVITLY 127

Query: 133 AYPQAVKRLSLSTLTKRSLLAGLFPFNSPAPLAFICITISLQL----SLTLFGQGS 188

+P V DLG FK SFL G+ P+ L I L++ +L LL LF S

Sbjct: 128 IFPTLVHFDLSYKQYFKISFL-LGVLPFR-TLLIMTILSLALLFLATFPILLFLP-AS 184

Query: 189 LLAIGISISLTYL 202

+A + + S + Y

Sbjct: 185 FVAALTWGFLFGY 198

An alignment of the GAS and GBS proteins is shown below:

Identities = 68/210 (32%), Positives = 117/210 (55%)

-368-

```

      :| |::: |:: | : :|:
VHPDWKGLLYVKSLLLSVAYLQYTLTVALTVALFLLAYLGLGVVFFSVSLISYCHMRIVYAVLLKVBQHGGEFQKRS
      150      160      170      180      190      200      210

```

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 290

A DNA sequence (GBSx0318) was identified in *S. galactiae* <SEQ ID 933> which encodes the amino acid sequence <SEQ ID 934>. Analysis of this protein sequence reveals the following:

```

10 Possible site: 51
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.1827(Affirmative) < succ>
15    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

20 >GP:AAC4392 GB:U43526 ORF-1 [Streptococcus pneumoniae]
    Identities = 48/151 (31%), Positives = 66/151 (42%), Gaps = 5/151 (3%)

    Query: 1 MIYDHLNLTHTYKDINPNLDLAIDYLLSHDLRLNDIGTYHISPEVILMVQSNQLES-PD 59
      MI + L Y +MP+ ID+L L NL G+ I + L++
25    Sbjct: 1 MIITKISLGTGYGVNPHFATLIDFLEKTGLNLTGSSIAIDGNRLFGNCFTYIADGQAG 60

    Query: 50 HIFSYHKYKYLDIHYVIGHEVIKLGKGDKEV-REY--LGDIGPIKCEETSFDLRDNYI 116
      FE R+KYLDIH V+E E + + + V V +EY DI E LR
    Sbjct: 61 AFFSTHQYKYLDIHLVLENEAMAVTSPENVSVTQEYDEEKDIELNYTGKVEQLVHLRAGEC 120

30    Query: 117 AFFFFPEAHQFMGMSLGRYVKKGVKLVMA 147
      FFE+ HQP + VKK V KV ++
    Sbjct: 121 LITFFDLHQFK-VRINDEPVKKVVFVKVAIS 150

```

No corresponding DNA sequence was identified in *S. pyogenes*.

- 35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 291

A DNA sequence (GBSx0319) was identified in *S. galactiae* <SEQ ID 935> which encodes the amino acid sequence <SEQ ID 936>. This protein is predicted to be sugar ABC transporter, permease protein (araQ).

- 40 Analysis of this protein sequence reveals the following:

```

    Possible site: 35
    >>> Seems to have a cleavable N-term signal seq.

    INTEGRAL Likelihood = -7.38 Transmembrane 245 - 261 ( 239 - 265)
    INTEGRAL Likelihood = -3.72 Transmembrane 140 - 156 ( 139 - 158)
45    INTEGRAL Likelihood = -3.61 Transmembrane 76 - 92 ( 71 - 94)
    INTEGRAL Likelihood = -2.81 Transmembrane 112 - 128 ( 107 - 128)
    INTEGRAL Likelihood = -1.59 Transmembrane 188 - 204 ( 186 - 204)

    ----- Final Results -----
50    bacterial membrane --- Certainty=0.3951(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

-369-

>GP:RAD35515 GB:AE001721 sugar ABC transporter, permease protein
 [Thermotoga maritima]
 Identities = 94/262 (35%), Positives = 158/262 (59%), Gaps = 1/262 (0%)

5 Query: 15 LILCLTLFLFIPFPYWMITGAFKSPQDTIIIPQWPKAPTLENFKALTVGNPALRWLM 74
 + + + V+F+ P + + + +FK + PP +PK P+LE + + + L +L N
 Sbjct: 9 IFIVFHLVFLVFLVFPVYAVSSFKMSILSYPPPTIFPKKSLGKYINVIKEYDLITLTYL 68

10 Query: 75 SVFISIMTFLVCTSSMAGYVLAKKFPYQKILFSLPIAAMALPKQVVLVFLVRIINFM 134
 ++F++ + + S M GY IAK +F+G + + S+F M + QV++VFL +I +
 Sbjct: 69 TLFVATVATVITFLVSVMTGGLARGKFWGILRPVNSMFTMTMFVSAQVINFLVFLVIRSL 128

Query: 135 GIHDTLWAVILPLVQWPFQVFLMKQPSRNIPTELLBSAKIDGCEIRTFINVAFFIVKPG 194
 G+ ++LW +I+P V P G+F+ Q+ ++IP ELLSSAKIDG E + F + FP+ KP
 15 Sbjct: 129 GLINSILWGLITPAVYITPGMFAVQYMKDIPDELLBSAKIDGANEWQIFWRIVFPLSKFL 188

Query: 195 FAALAIFTTINTWYDFMQLWILTSRNILTSIGVATMQAEN-ATNYGLIMGAALAAVP 253
 AALAIF+P NND+ + L+++ RN T+ L +AT+Q E + I+A + L +P
 Sbjct: 189 VAALAIFSTFTNRNDFVLPLVNVNRNLYTLQLALATIQEYGGABMTLAFSLITLIP 248

20 Query: 254 IVTVFLVQKSFQGTGIMGAVK 275
 + +FL+EQ+ F +GI G +K
 Sbjct: 249 TLITFLFLQRLFKGIMAGGLK 270

- 25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 937> which encodes the amino acid sequence <SEQ ID 938>. Analysis of this protein sequence reveals the following:

Possible site: 40
 >>> Seems to have a cleavable N-term signal seq.

30 INTEGRAL Likelihood = -6.37 Transmembrane 245 - 261 (240 - 265)
 INTEGRAL Likelihood = -5.15 Transmembrane 140 - 156 (139 - 158)
 INTEGRAL Likelihood = -2.97 Transmembrane 111 - 127 (107 - 128)
 INTEGRAL Likelihood = -2.87 Transmembrane 76 - 92 (75 - 93)
 INTEGRAL Likelihood = -1.59 Transmembrane 188 - 204 (186 - 204)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.3546(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 40 The protein has homology with the following sequences in the databases:

>GP:CAB59597 GB:AL132662 probable sugar transport inner membrane
 protein [Streptomyces coelicolor A3(2)]
 Identities = 88/262 (33%), Positives = 147/262 (55%)

45 Query: 15 VMLCVLTILFIPFPYWMITGAPKAQNDTIMIPQWNPQAPTLENFKALTVGNPALRWLM 74
 ++L L ++P P W++ + + A+ PP KP + ++ + +N N
 Sbjct: 38 LLLAPLALVFAVFLVWLVLSSVMSNARINRPPALPKSGIDLGGRYVLGNAMPFRPFIN 97

50 Query: 75 SVFISVATMFLVCTSSSLAGYALAKKRFVQGRILFSLPIAAMALPKQVVLVFLVRIINFM 134
 S+ +S T+ SLAGYA A+ RF G R+L + +A MA+P Q+ +P ++ +
 Sbjct: 98 SLIVSAVTVAANLVFQSLAGYAFARMRFAGSRVIMSLMLATNAVPPQLTMTFTFLVWKL 157

55 Query: 135 GIHDTLWAVILPLVQWPFQVFLMKQPSRNIPTELLBSAKIDGCEIRTFINVAFFIVKPG 194
 G+ DTL I+I+P + PF VFL+QF +PF EL E+A IDGC +R + + P+ +P
 Sbjct: 158 GLIDTILGALIVPSLVTTPFAVFLRQFLGLPRELSKAWIDOCRLRVLMRIVLPERPA 217

Query: 195 FAALAIFTTINTWYDFMQLWILTSRNILTSIGVATMQAENATNYGLIMGAALAAVPI 254
 A +A+ TF+ TWND L+ + T+ LG+ T Q+ T + +MDG + +P+
 Sbjct: 218 LATVAULTLFTTWDLTWPLIAINHCTYTTLTGLGITTFQGGHTQAAVNAVGVITVLEV 277

60 Query: 255 VTFLVFLVQKSFQGTGIMGAVK 276
 + FL QK+F Q IT +KG
 Sbjct: 278 LLAFLGAQKTFIQSITSSGLK 299

- 65 An alignment of the GAS and GBS proteins is shown below:

-370-

Identities = 245/276 (88%), Positives = 262/276 (94%)

Query: 1 MKKQTFSAYNFLTALILCLITVLGIFPPFWINTGAFKQSDTIIIPQWMPKADTLENFK 60
 M KK +A +LT ++LC+LT+LGFPPFWINTGAFK+Q DTI+IPQWMPKAPT+ENFK
 5 Sbjct: 1 MTKKKLTASDILT+VMLCLITLIFLGFPPFWINTGAFKQADTIIIPQWMPKAPTLENFK 60

Query: 61 ALTVQNPALRWLWNSVFSIMIMFLVCCSTSSMAQYVLAKKRFTGQKILFSLPIAAMALPK 120
 AL VQNPAL+WLWNSVFSIS+ TMFLVC TSS+AGY LAKKRFTGQ++LFS+PIAAMALPK
 10 Sbjct: 61 ALVQNPALRWLWNSVFSIVATMFLVCGTSSLAGVALAKKRFTGQKILFSLPIAAMALPK 120

Query: 121 QVVVLPLVRIINPMGIHDTLWAVILFVAGWPFQVFLMKQPSENIPTELLESAKIDGCGEI 180
 QVVVLPLVRI+NPNGIHDTI, AVILFVAGWPFQVFLMKQPSENIPTELLESAKIDGCGEI
 15 Sbjct: 121 QVVVLPLVRIVNPNGIHDTLAAVILFVAGWPFQVFLMKQPSENIPTELLESAKIDGCGEI 180

Query: 161 RTFINVAFPIVKPGFAALAIPTFINVNDYFMGLVMTLSKNNLTISLGVATMQAEMATNY 240
 RTF NVAFPIVKPGFAALAIPTFINVNDYFMGLVMTLSR NLTISLGVATMQAEMATNY
 20 Sbjct: 161 RTFINVAFPIVKPGFAALAIPTFINVNDYFMGLVMTLSRENLTISLGVATMQAEMATNY 240

Query: 241 GLINAGRAA+RAVPIVTVFLVFKQSFTQGITMGAVG 276
 GLINAGRAA+RAVPIVTVFLVFKQSFTQGITMGAVG
 25 Sbjct: 241 GLINAGRAA+RAVPIVTVFLVFKQSFTQGITMGAVG 276

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 292

A DNA sequence (GBSx0320) was identified in *S. agalactiae* <SEQ ID 939> which encodes the amino acid sequence <SEQ ID 940>. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have a cleavable N-term signal seq.
 30 INTEGRAL Likelihood = -10.83 Transmembrane 74 - 90 (64 - 96)
 INTEGRAL Likelihood = -6.37 Transmembrane 108 - 124 (107 - 126)
 INTEGRAL Likelihood = -5.84 Transmembrane 270 - 286 (265 - 290)
 INTEGRAL Likelihood = -5.20 Transmembrane 161 - 177 (156 - 182)
 35 INTEGRAL Likelihood = -0.16 Transmembrane 219 - 235 (219 - 235)

----- Final Results -----
 bacterial membrane --- Certainty=0.5331(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 40 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB05584 GB:AP001513 sugar transport system (permease) (binding
 protein dependent transporter) [Bacillus halodurans]
 Identities = 106/289 (36%), Positives = 168/289 (57%), Gaps = 6/289 (2%)
 45 Query: 9 REIMIAVAPLAPILLPLFLIPAPWMMGVSFFMYSM-TQPTIGLANYNMF-HDSIF 66
 +E Y P+AP ++ P IF EM+ SE ++ + + G NY R+P D +F
 Sbjct: 25 KEYFWKYLFIAPPIIGFAIFALGPMILSYIVSDIFDLYNEPWTGADNYRLVFTDDEL 84

Query: 67 MGLSINTLVIVIGSVPPVVPFSLVAANTYKKAFFSRSPYRCVFELPVVVGSAVTVVWK 126
 K+ + N+ +G +P+ + SL +A +K V + +R FFLP V+ +D+T++H+
 50 Sbjct: 85 RKTVPNTFYAALG-IPIGMAVSLGIAVAINQK-VKGLAFRTAPFLPAVSSVVAITLLMR 142

Query: 127 WYDPMSSGILNVLKSGHIVIRQISLWGDKRWALLAIIILLTSSWGQPIILYIARMCNI 186
 WI++ G+LN +L +V WL D+ W+ A+II + +G +IIX+AA+ +
 55 Sbjct: 143 WIFNADFGLLNIMLN--YVGIIKGOWLSEKRWPMIIGVWGGGLNIMILYLAALQGV 200

Query: 187 DNSLCEARVDGANEMQVFWQIKWPELLETFLYIAVITINSFOCALIQLITSGDNTYS 246
 + +L EAA +DG N Q F I P6+ PTT +I + +TI + Q F + +T GGNNTYS
 60 Sbjct: 201 N8ALYEADIDGNAWQKFIHTVPSISPTTFPILITSTIGCALQDQRFMINTEGCPNYS 260

Query: 247 TSTIAYLYEKAPKLSYGVANTMGVFLAVMIALISPAQPKILGNVSY 295

-371-

T+T++YYL+ AP+ E GYA+ M L ++I +I+ PK+ V Y
 Sbjct: 261 TITVVVYVFLAFAFRYMEGYASAMAVLGIILLITITINFKIAKQVHY 309

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 941> which encodes the amino acid sequence <SEQ ID 942>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -12.74	Transmembrane	55 - 71 (44 - 78)
INTEGRAL	Likelihood = -10.83	Transmembrane	109 - 125 (98 - 130)
INTEGRAL	Likelihood = -6.21	Transmembrane	304 - 320 (299 - 324)
INTEGRAL	Likelihood = -6.00	Transmembrane	142 - 158 (141 - 160)
INTEGRAL	Likelihood = -5.04	Transmembrane	196 - 212 (190 - 216)
INTEGRAL	Likelihood = -0.16	Transmembrane	253 - 269 (253 - 269)

----- Final Results -----
 bacterial membrane --- Certainty=0.6095(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAB05584 GB:AP001513 sugar transport system (permease) (binding protein dependent transporter) [Bacillus halodurans]
 Identities = 113/310 (36%), Positives = 176/310 (56%), Gaps = 9/310 (2%)

Query: 25 KVEQKEVFQVNNVKLMR---STLISYAFAPLVLFVFLIPIMMGFVTSFFNYSM- 80
 +VE +E K K R E Y F+AP ++ P I F L PM+ SF ++ +
 Sbjct: 4 EVETPRETKTKARKQKRRLNKEYFGYLFIAPIIGFAIFALGPMLYSIVSFTDOLY 63

Query: 81 TRFTVFGFANYARMF-QDPIPMKSLINTLIIIVIGSVPVVFFSLFVAAKTYDKNVARSF 139
 E + G MY R+F D +P K+ + NT +G +P+ + SL +A K V +
 Sbjct: 64 NEPVWIGADNYRLFVTDLFRKTVENTFYAALG-IPIGMAVSLGIAVALNQK-VKGIAL 121

Query: 140 YRAVFFLPVVTGSVAVTVMWKIYDPMGSLINYLKARHVEQNIISWLGDKHALLAIV 199
 +R FFLP V+ VA+T++W+II++ G+LN +L Y + WL D+ WA+ A+I+
 Sbjct: 122 ERTAFPLPAVSSVAITLNRWIFNFIHAFGLLIMLNTYVGI--HGPFNSDEKHAMPAII 179

Query: 200 ILLTTSVGQPIILLYIAAMGNIDNSLVEAARVDGATEQVFWNIKWPSSLPTTLYIAVITT 259
 + +G +ILY+AA+ ++ +L EAA +DG +Q F+I PS+ PTT +I +T
 Sbjct: 180 QGVWGLGIMLILYLAALQGVNPALEYAADIDGNAWQRFIHITVESIPITFFILITST 239

Query: 260 INSEQCFLIQLLTSGGFNYSTSTLNYYLYEKAFLSEYGYANTMGVFLAVIAIISPAQ 319
 I + Q F ++T GGFNYST+T++YYL+ AP+ E GYA+ M L ++I II+
 Sbjct: 240 IGAIGDQRFPMITSGGFNYSTTTT+VYVFLAFAFRYMEGYASAMAVLGIILLITITIN 299

Query: 320 PKILGNVEY 329
 PK+ V Y
 Sbjct: 300 PKIAKQVHY 309

An alignment of the GAS and GBS proteins is shown below:

Identities = 263/295 (89%), Positives = 278/295 (94%)

Query: 1 MRTNKLAKNREIMIAVAFAPILLFSLIFVFAPMVGFVTSFFNYSMITQFIGLANRIM 60
 + NKLNRRET-I+YAFAPL+L+FP+IFV PM+MGFVTSFFNYSMIT+FTP+G ANY RM
 Sbjct: 35 VNNVKNLAKNRETLISYAFAPLVLFVFLIPIMMGFVTSFFNYSMITQFIGLANRIM 94

Query: 61 FHDSPMKSLINTVILIVIGSVPVVFFSLFVAANTYKKNVRSFTRCFVFLPVVTGSVA 120
 F D I PMKSLINT+IIVIGSVPVVFFSLFVA TY+KNV +RSFR VETLFPVVTGSVA
 Sbjct: 95 PQDPIPMKSLINTLIIIVIGSVPVVFFSLFVAAKTYDKNVARSFTRAVFPLPVVTGSVA 154

Query: 121 VTVWKKIYDPMGSLINYLKSGHIVBQNIISWLGDKHALLAIIILLTTSVGQPIILYI 180
 VTVWKKIYDPMGSLINYLK HVBQNIISWLGDKHALLAII+ILLTTSVGQPIILYI
 Sbjct: 155 VTVWKKIYDPMGSLINYLKARHVEQNIISWLGDKHALLAII+ILLTTSVGQPIILYI 214

Query: 181 AAMGNIDNSLCEAARVDGANEQVFWQIKWPSLLPTTLYIAVITTSNQCFLIQLLT 240

-372-

AAMGNIINSL KAARVDGA E QVFM IKWPSLLPTLYIAVITTTNSFQCQALQLLTS
 Sbjct: 215 AAMGNIINSLKAARVDGATEFQVFWNIKWESLLPTLYIAVITTTNSFQCQALQLLTS 274

Query: 241 GGPYISTSLMYL.YEKAFKLSGTYANTMGVFLAVMIALISFAQFKILGNDEV 295
 GGPYISTSLMYL.YEKAFKLSGTYANTMGVFLAVMIA+ISFAQFKILGNDEV
 Sbjct: 275 GGPYISTSLMYL.YEKAFKLSGTYANTMGVFLAVMIALISFAQFKILGNDEV 329

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 293

A DNA sequence (GBSx0321) was identified in *S.agalactiae* <SEQ ID 943> which encodes the amino acid sequence <SEQ ID 944>. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12516 GB:Z99107 similar to sugar-binding protein [Bacillus subtilis]
 Identities = 54/187 (28%), Positives = 90/187 (47%), Gaps = 14/187 (7%)

Query: 19 MFACVDSSQSVMAAEKD-KVEITWMAFPTTQEKAKDVGTYEKKVIFAKERKIPNIKVK 77
 MF+ + + ++D + I WW + D Y KVI+ +ERKIP+ ++
 Sbjct: 1 MFSGCSAGEASGKKEDVTLRIAMWG----GQPHMD---YTKVIELYERKIPHVHIE 51

Query: 78 LETIDFTSGPEKITTAIEAGTAPDLFLDAPGRIQYQKNGKLADLNDLTDQIFKDN-- 135
 E ++ ++K+ AG EDV+ + QYK +L DL D I DV+
 Sbjct: 52 AEPANWDDYWKLAPMSAGQLFDVIMQDITAYLAQYKKNQLEDLPTTKDGTI-DVSSI 110

Query: 136 NQNIQASKSGDEANNYPISAPFMAFNKMLDAGVLKLVKEGWTTSDPEKVLKALGN 195
 ++N+ K +K Y + + N+ +LK AGV + +E WT D+EK+ L+
 Sbjct: 111 DENMLSGKIDNKLYGFTLGNNVLSVIANEDLLKAGV-SINQENNTWEDYKLAIDLQ 169

Query: 196 KGYTPGS 202
 K GS
 Sbjct: 170 KAGVGS 176

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 945> which encodes the amino acid sequence <SEQ ID 946>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> May be a lipoprotein

----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

!GB:Z99107 similar to sugar-binding protein [Bacillu... 82 2e-14

>GP:CAB12516 GB:Z99107 similar to sugar-binding protein [Bacillus subtilis]
 Identities = 105/446 (23%), Positives = 176/446 (38%), Gaps = 71/446 (15%)

Query: 24 GKSQKEZGASKSIDAKTRITWAFVPTQSKAEDVGTYEKKLIAAPEKANPRIVKLET 83
 G S E + K + I WW + D Y K+I +EK NP + ++ E

-373-

Sbjct: 4 GCSAGEASGKKEDVTLRIAWWG-----GQPRHD---YTTKVIKLYSKKNPHVHEAEF 54

Query: 84 IDPTSGPEKITTAEAGTAPDVLFDARGRIIQYKGNKLADNLDLPTBPTKDVN--NDK 141
 ++ +K+ AG PDV+ + QYK +L DL +T++ T DV+ ++

5 Sbjct: 55 ANWDDYKKKLAPMEASAGGLPDVLIQMLTAYLACQYKGNQLRLDLP-YTKDGTIDVSSIDEN 113

Query: 142 LIQASKAGDTAYMYPISAPFYMALNNKMLKMDGVLDLVKEGWTDDPRFKVILKALKK-- 199
 ++ K + Y + + + N+ +LK AGV + +E WT +D+EK+ L+K+

10 Sbjct: 114 MLGGKIDNKLGYPTGLVGNVLSVIANEDLLKAGV-SINQENWNTWSDYBLKLAYLQGEKAG 172

Query: 200 -----GTNPGSFFANGQGGDQGFAPFANLYSSHITDDKV-----TKYTT 239
 G +P F +G R + + DD++ T T

Sbjct: 173 VYGSNGKHPDIFFFPYLRTKGERFYKEDGTGLAYQDQLFVDYFERQLRLVKAKTSPT 232

15 Query: 240 DNASIKAMTKISNWKIDGLMNGSQYDGSADIQNFANGQTSFTILWAPAQPGIQAALL 299
 D++ IK M +D ++ G SA N++N F +A+L +

Sbjct: 233 DESAQIKM-----EDDFIVGK-----SAITWYSGNYLGF-----ARLTD 269

Query: 300 ASKVLYLEIPFPDDGKPELEVLVNGFAVNNKDEQKVAASKTFIQIADDKKQKPNV 359
 + YL P + L + E K A+K FI F +++E + +

20 Sbjct: 270 SPLSLYLF---PEQMKEKALTLPKSMPLFPIKSESHKGEAK-FINFPVNNE-ANQLIK 324

Query: 360 RTGAPFVRTSYGDLYKDKRMK---IAWTKFYSPYNTIDG---GFAEMRTLWFPVQ 411
 PV D K K E + I E + S + D G AE+ L +

25 Sbjct: 325 GERGVVPSDKVADAIPKLANEETNIVEYVETASINISKADPFEPVGSAEVILKLTSD 384

Query: 412 AVENKDEKPEALKAFTKANKTIKK 437
 + DE A K P +KAM+ ++

30 Sbjct: 385 QILYQVSEKAATKPKIAMEILR 410

An alignment of the GAS and GBS proteins is shown below:

Identities = 352/438 (80%), Positives = 384/438 (87%), Gaps = 4/438 (0%)

Query: 1 MSIKSVIGFCLGAALSMFACVDSQSVMAAEKD---KVEITWMAFFTQIEKALDGVG 57
 M++KK LGA+ L + AC SQ A K K EITWMAFF FIQERA+DGVG

35 Sbjct: 1 MNMKLASLAMLGASVLGLAACGKSGKERGASKSDTAKTEITWMAFFVFTQIEKADGVG 60

Query: 58 TYEKVKIAFEKIQNPHIKVLETDFTSGPEKITTAEAGTAPDVLFDARGRIIQYKGN 117
 TYEKK+I AFEX NP IKVLETDFTSGPEKITTAEAGTAPDVLFDARGRIIQYKGN

40 Sbjct: 61 TYEKKLIAFEKANPEIKVLETDFTSGPEKITTAEAGTAPDVLFDARGRIIQYKGN 120

Query: 118 KLADNLDLPTDQIKDVNNKHIIQASKSGDKAWNYPISAPFYMAFNKMLKMDAGVLKL 177
 KLADNLDLPT+P KDVEN +IQASK+D ANYPISAPFYMA NKMLKMDAGVL LV

45 Sbjct: 121 KLADNLDLPTKEPTKDVNNKLIQASKAGDTAYMYPISAPFYMALNNKMLKMDAGVL 180

Query: 178 KEGWTTSDPEKVLKALKHKGYPGSGFFANGQGGDQGFAPFANLYSAPITDKEVTKYTTD 237
 KEGWTT DPEKVLKALK+KGY PGSGFFANGQGGDQGFAPFANLYS+ ITD +VTKYTTD

Sbjct: 181 KEGWTTDDPEKVLKALKDKYNGPSSFFANGQGGDQGFAPFANLYSSHITDDKVTKYTTD 240

50 Query: 238 TMSVKSMMKIVEMIKKGYLMNGSQYDGSADIQNFANGQTSFTILWAPAQPGIQAALL 297
 NS+K+M KI WIK G +MNGSQYDGSADIQNFANGQTSFTILWAPAQ QAILEA

Sbjct: 241 DNASIKAMTKISNWKIDGLMNGSQYDGSADIQNFANGQTSFTILWAPAQPGIQAALL 300

Query: 298 SKVDYLEVSPFSDGKPDLEYLVNGFAVNNKDEKVKASKEPTFIADDKKQKPNVDTR 357
 SKVDYLE+PFS+DKPD+LEYLVNGFAVNNKDE KV ASK PI FIADDK+MKPE+V+R

55 Sbjct: 301 SKVDYLEIPIFPDDGKPELEVLVNGFAVNNKDESKVAASKTFIQIADDKKQKPNVVR 360

Query: 358 TGAFPVRTSYGDLYKDKRMKIKSGWTKFYSPYNTIDGFAEMRTLWFPVQSVNGDEK 417
 TGAFPVRTS+GDLYK DKRM KI++WT+YSPYNTIDGFAEMRTLWFPVQSVNGDEK

60 Sbjct: 361 TGAFPVRTSYGDLYK-DKSMKIASWTKFYSPYNTIDGFAEMRTLWFPVQSVNGDEK 419

Query: 418 PADALKDFTQKANDTIKK 435
 P DALK FT+KAM TIKK

65 Sbjct: 420 PEDALKAFTEKANKTIKK 437

SEQ ID 944 (GBS16) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 3 (lane 9; MW 49kDa).

The GBS16-His fusion product was purified (Figure 92A; see also Figure 189, lane 9) and used to immunise mice (lane 1 + 2 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 92B), FACS (Figure 92C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 294

A DNA sequence (GBSx0322) was identified in *S. agalactiae* <SEQ ID 947> which encodes the amino acid sequence <SEQ ID 948>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 945> which encodes amino acid sequence <SEQ ID 946> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AA056999 GB:AR001166 conserved hypothetical protein [Borrelia
burgdorferi]
Identities = 107/225 (47%), Positives = 147/225 (64%), Gaps = 6/225 (2%)

Query: 12 QIKNGIIVSCQALPGEPLYTESGGVMPILALAAQKAGAVGIRANSVRDIKEIQEVNLP 71
+IK G+IVSCQAL EPL+ S +M +ALAA+ GA+GIRAN V DI +I+ +LPI
Sbjct: 6 KIKRGLIVSCQALENEPLH--SSFIMSKQALAAKIGGAIGIRANGVNDISQIKLEVDLPI 63

Query: 72 IGIKIKREYPPQEPFITAIMTEVDQLASLDIAVIALDCTLRERHGLSVVEFTQIKIRKYP 131
IGIIK+ Y + FIT TM E+D+L + + +IALD T R R DG+ + +F + IK+KYP
Sbjct: 64 IGIKIKYNNKCDVFTPTMKEDISLNEGVDIHALDATFRNRDPGVLLDDFFRNKIKKYP 123

Query: 132 EQLLQADISTFEEDGKQAFEGAVDVFVGTTLISGYTOYSR--QREGPDIELLNKLQAGI--D 187
+Q LQADIS+ +E NA + G DF+GTTL GYT + D L L + +
Sbjct: 124 KQCLQADISLDEALNADKLQDFDFTLTYGYTKNTINGLIALNDENFLRLTLNLSNKEST 183

Query: 188 VIAEGKIHTPKQANEINHIGVAGIIVVGGAITRPKEIARFISGLS 232
+I EGKI TP +A + +GV +VVGGAITRP EI ++F+ ++
Sbjct: 184 LIVEGKIDITPLKAKCFEMGVDLVVVGGAITRPKEITKKFVEKIN 228
```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 949> which encodes the amino acid sequence <SEQ ID 950>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -1.49 Transmembrane 175 - 191 ( 175 - 192)

----- Final Results -----
bacterial membrane --- Certainty=0.1595 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

-376-

The protein has homology with the following sequences in the databases:

>GP:AA28762 GB:AF130859 putative N-acetylmannosamine-6-P epimerase
[Clostridium perfringens]
Identities = 113/225 (50%), Positives = 148/225 (65%), Gaps = 5/225 (2%)

5 Query: 10 LMBQLKGGIIVSCQALGPEFLYSEGTGIMPLAKAKAQAGAVGIRANSVRDIKEIQAITD 69
+++ +KG +IVSCQAL EPL+S IM MR AA++ GA IRA + DI EI+ +T
Sbjct: 1 MLFVVGKLNIVSCQALSDEPLHSSF--IMGRMAIAAKQQAARAIQAGIDIDINEIKVTK 58

10 Query: 70 LPIIGIIRKYPQPFITATMTEVDQLAALNIAVIAIDCTKRDRHDLGDIASIRQVKE 129
LPIIGIIRK++Y E +IT TM EVD+L + +I +D TKR R +G +I + +
Sbjct: 59 LPIIGIIRKRYDSELEYITPMKEVDLELLKTDCEMIGLDAIKRKRNGENIKDLVDIAHA 118

15 Query: 130 KYPNQLMADISTFDBGLVAHQAGIDFVGTTLGGITPYRQEGPQVALIEALCK-AGIA 188
K +L MADIST +EG+ A + G D V TTLGGITPYR+Q D L+E L K I
Sbjct: 119 K--GRLMADISTLEBGLIABKLGDFCVSTTLGGITPYRQNSVDPELLESLVTKIP 176

Query: 189 VIAEGKIHSPEEAKKINDLGAVAGIVVGGAITRPFKEIARPFIALK 233
VI EG+L++PSE KK DLS VVGGAITR++I +RF + LK
20 Sbjct: 177 VICEGRINTPEELKKALDLGAVAVVGGAITRPFQITKRPFDILK 221

An alignment of the GAS and GBS proteins is shown below:

Identities = 172/227 (75%), Positives = 202/227 (88%)

25 Query: 5 SKEAFKKQIKNGIIVSCQALGPEFLYSEGGVMPILALAAQAGAVGIRANSVRDIKEIQ 64
+KE +Q+K GIIVSCQALGPEFLY+E+GG+MPL+A AAEAGAVGIRANSVRDIKEIQ
Sbjct: 6 TKEKMEQLKGGIIVSCQALGPEFLYSEGTGIMPLAKAKAQAGAVGIRANSVRDIKEIQ 65

30 Query: 65 EVTNLPILIGIKREYPPQPFITATMTEVDQLASLIDIAVIALDCTLRSDHGLSVVEFIQ 124
+T+LPIIGIIRK++YPPQPFITATMTEVDQLA+L+LAVIA+DCT R+RDGL + FI+
Sbjct: 66 AITDLPIIGIIRKRYDSELEYITPMKEVDLELLKTDCEMIGLDAIKRKRNGENIKDLVDIAHA 125

Query: 125 KIKRKYPBQLMADISTFEEGNAFEAGVDFVGTTLGGITPYRQEGPDIENLKLQCA 184
++K KYP QLLMADISTF+EG A +AG+DFVGTTLGGIT PYRQE GPD+ L+ LC+A
35 Sbjct: 126 QVKEKYPNQLMADISTFDBGLVAHQAGIDFVGTTLGGITPYRQEGPDIENLKLQCA 185

Query: 185 GIDVIAEGKIHSPEEAKKINDLGAVAGIVVGGAITRPFKEIARPFIAL 231
GI VIAEGKIHSPEEAKKINDLGAVAGIVVGGAITRPFKEIARPFIAL L
40 Sbjct: 186 GLAVIAEGKIHSPEEAKKINDLGAVAGIVVGGAITRPFKEIARPFIAL 232

Based on this analysis, it was predicted that these proteins could be useful antigens for vaccines or diagnostics.

Example 295

A DNA sequence (GBSx0323) was identified in *S. galactiae* <SEQ ID 951> which encodes the amino acid sequence <SEQ ID 952>. This protein is predicted to be group B streptococcal surface immunogenic protein. Analysis of this protein sequence reveals the following:

Possible site: 25
>>> Seems to have a cleavable N-term signal seq.

50 ----- Final Results -----
bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

55 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 953> which encodes the amino acid sequence <SEQ ID 954>. Analysis of this protein sequence reveals the following:

Possible site: 25

-377-

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

5 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 182/437 (41%), Positives = 240/437 (54%), Gaps = 53/437 (12%)

10 Query: 1 MGNKKVLLTSTAASLLSVASVQAQETDTTWIARTVSEVKADLVKQDNSESYTVKYGDT 60
 M + KK L +++A SL+ +A+ QAQE WT R+V+R+K++LV DN +YTVKYGDT
 Sbjct: 1 MIITKSLFVTSVALSLVPLATAQAQE----WTPRSVTIKSELVLVDNVPITYTVKYGDT 56

15 Query: 61 LSVISEAMSIDMNLAKINNIADINLIYPETTLITVYDQKSHIATSMKISTPATNAAGQT 120
 LS I+EAM ID++VL IN+IA+I+LI+P+T LT Y+Q AT++ ++ PA++ A +
 Sbjct: 57 LSTIARAGMIDVHVLGDINHIANIDLIPDPTILTANYNQHQQ-ATNLIVQAFASSPASVS 115

20 Query: 121 TATVLDKINQSVADQKVLNITISBGMTP-EAATTIVSPMKTYSSAPALEKEVLAQBQA 179
 Q S Q ++ TP + TT + K S S A S E+ +
 Sbjct: 116 HVPSPSEFLQAGATSQPTV--PMAPPATPSDVPTTFFASAKPDSSVTA--SSELTSSTND 171

25 Query: 180 VSGAANNDQVSPAPVKISITSEVPAAKEEVKPTQTSVQSQTIVSPASVAETPAPVAKVAP 239
 VS ++E V P A E T V T +S A +A P P +
 Sbjct: 172 VSTELSSRSQKQPEVPQAEVPTPKAAE----TTEVEPKTDISEAFTSANRFPVNEASGE 226

30 Query: 240 VRTVAAPRVAIVKVTTPKVTGTASPERVSAFAPV---VITTSPTATSKLQATEVKSVPVA 296
 + AAP + A E SAPA TTS AT + L
 Sbjct: 227 EVSSAAP-----AQAFAKEETSAPAAQKAVADTTSVATSNGL----- 264

35 Query: 297 QKAPTATFVAQPASTTNAAHPENAGLOPHVAAYKEKVASTYGVNEFSTYRAGDPGDHG 356
 AP A + P RAGLOP AA+KE+VAS +G+ FS YR GDGPDHG
 Sbjct: 265 SYAPNH-----AYNFMKAGLOPQTAAPKEE+ASAFTISYSGYRGGDPGDHG 311

40 Query: 357 KGLAVIDFVINGTQALGNKVAQSTQNRANINISYIVNQQKFYSNTNISYGFANTWRNMPD 416
 KGLA+DF+V N ALG++VAQY+ +MA ISYIVN+Q+FY+ SIYGA. TNN MPD
 Sbjct: 312 KGLAIDFMVPENSALGDQAQAIDHMAERGISYIVNQRFTAPFASISYGFATYNNMPD 371

40 Query: 417 RGGVTANEDYDHHVHVSFN 433
 RG +T NEDYDHHVHVSFN
 Sbjct: 372 RGSITENYDHHVHVSFN 388

A related GBS gene <SEQ ID 8539> and protein <SEQ ID 8540> were also identified. Analysis of this protein sequence reveals the following:

45 Ldcp: Possible site: -1 Crend: 3
 SRCFLG: 0
 McG: Length of UR: 20
 Peak Value of UR: 1.96
 Net Charge of CR: 2

50 McG: Discrim Score: 2.95
 GrH: Signal Score (-7.5): 3.84
 Possible site: 23

55 >>> Seems to have a cleavable N-term signal seq.
 Amino Acid Composition: calculated from 24
 ALOM program count: 0 value: 4.29 threshold: 0.0
 PERIPHERAL Likelihood = 4.29 58
 modified ALOM score: -1.36

60 *** Reasoning Step: 3

 Rule gpol

----- Final Results -----

65 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

-378-

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

SEQ ID 8540 (GBS322) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 9; MW 52kDa). The GBS322-His fusion product was purified (Figure 214, lane 10) and used to immunise mice. The resulting antiserum was used for FACS (Figure 267), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 296

A DNA sequence (GBSx0324) was identified in *S.galactiae* <SEQ ID 955> which encodes the amino acid sequence <SEQ ID 956>. Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -1.86 Transmembrane 5 - 21 (4 - 21)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1744(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AA046072 GB:U50357 zocin A endopeptidase [Streptococcus
 zooepidemicus]
 Identities = 163/274 (59%), Positives = 196/274 (71%), Gaps = 11/274 (4%)
 Query: 25 VLADTYVRPIDNRIIT:GFGNGYFGHGVYAVPTGTIIRAVADGTVKFAAGAGANFSWTD 84
 V A T Y R P + D G I T T G F N G Y P G H G V D Y A V P G T + R A V A + G T V K F A G A N W N
 Sbjct: 21 VSAATYTRFLDTGNITITGFGNGYFGHGVYAVPTGTIIRAVADGTVKFAAGAGANFSWTD 80
 Query: 85 LAGNCVMIQHADGMHSGYAHMSRVVARTGEKVKQGDIIYVGATGMATGPHLHFEFLPAN 144
 +AGNCV+IQHADGMH+GYAH+S++ T VKQG IIGY GATG TGPHLHFE LPAN
 Sbjct: 81 MAGNCVLIQHADGMHSGYAHLSKISVSTDSVTKQGGIIGYTGATGQVTPGPHLHFEFLPAN 140
 Query: 145 PNFQNGFPHGRINPTSLIANVATFSGKTKASAPSIKPLQSAFVQNSKIAKVRVDELQKV 204
 PN+QNGF GRI+PT IAN F+G T + F N LK+Y+VD+LQK+
 Sbjct: 141 PNMQNGFSGRIDPTGTYLADPVFNGTTFTE-----PTTPTIN---LKIYKVDLQKI 189
 Query: 205 NGVNLVKNLTPTGTFWMDNGIBSEIDEVDANGNLTDQVLQKGGYFIPNPKTKIKTVE 264
 NG+W V+NN L PT F W DNGI A ++ EV +NG T+DQVLQKGGYF+NP +K+V
 Sbjct: 190 NGIYQVRNLIPTDPTFVMDNGIAADVIEVTSNGITRTSDQVLQKGGYFIPNPNKNSKVG 249
 Query: 265 KPIQTACGLTWAKTRFANGSVMNLRVDSQELLY 298
 P++G+ GL+VA+ F G +VNL + LLY
 Sbjct: 250 TPKKSGGLSQAQVNFPTTGNNVNLITTSKINILY 283

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8541> and protein <SEQ ID 8542> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 6
 MoG: Discrim Score: 6.63
 Gvi: Signal Score (-7.5): -2.97
 Possible site: 23
 >>> Seems to have an uncleavable N-term signal seq
 ALON program count: 1 value: -1.86 threshold: 0.0
 INTEGRAL Likelihood = -1.86 Transmembrane 5 - 21 (4 - 21)
 PERIPHERAL Likelihood = 5.57 50

-380-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2815 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04352 GB:AP001509 phosphoribosylaminoimidazolecarboxamide
 formyltransferase/IMP cyclohydrolase [Bacillus halodurans]
 Identities = 310/515 (60%), Positives = 390/515 (75%), Gaps = 4/515 (0%)

10

Query: 1 MTKRALISVDSKGIIDFAKELKNLGNIIISTGGTKVALDDAGVETIADVDVTFPPE@D 60
 M +RAL+SVS+K GI+ FAK L +I+STGGTK AL +AG+ I DVTGFPPE+D
 Sbjct: 1 MKRRALISVSNKGIIVPFAKALVGEVEIIVSTGGTKRALQKAGI PVTGISDVTGPIAELD 60

15

Query: 61 GRVKTLPNHNHOGLLARRDDSHLQAAXNNIELIDLVVNNLYPFKEITILRPDVTYDLAV 120
 GRVKTLPNHNHOGLLA R+ D HL +++I ID VVNNLYPF+TI +P+ T+ A+
 Sbjct: 61 GRVKTLPNHNHOGLLAMRERDEHLAQLNPHHRIPIIDFVVNNIYPQQTIAKPEATFADAI 120

20

Query: 121 ENIDIGGFSMLRSAAKSHASVTVVDSADYATVLGELACASCTTFTKTRQRIAAKAFRHTA 180
 ENIDIGGFSMLR+AAKSH VTVVVD DY TVL ELAD +T+RIAAK FRHTA
 Sbjct: 121 ENIDIGGFSMLRAAKSHQHVTVVVDPVYETVLKELADQGNVATETKRRFAAKVFRHTA 180

25

Query: 181 AYDALIAEYFTAQGEAKPEKLITITLYLKQAMRYGENPQQDADFYQKALFDYDYSIASAQ 240
 AYDA+IAEY T VGE PE LT+T++ KQ +RYGENP Q A FYQK L SIA ARQ
 Sbjct: 181 AYDAIAEYLTDAVGEESPESLTVTFEKKQLRYGENPHQKATFYQKLGAKASIAARQ 240

30

Query: 241 LNKELSFMINIRDADAARIITRDFKDSPTVVALKHMNPOGIGQADDISTAMDYAEADPV 300
 L+KELSF+MINI DADA+ I+++FK+ P VA+KHMNPOG+G +I+ A+D AYEADEV
 Sbjct: 241 LNKELSFYNNINDADAALSIVKEFKE-PAAVAVKHMNPOGVGTGETIKAPDKATYAEDEV 299

35

Query: 361 ASEVEAEYTVGGVLLVQNNQVVAENPSCMVVTRDQPTQEARALEFAWAKIYKVSNG 420
 ++ E T + GS LVQ +D ++ ++ T R+PTE E AL+ AW+ +K+VKSIV
 Sbjct: 358 ~NQAEKRITSINGSLVQEDTYGFEASIKIPTYREPTAEWELKLANRVVVKVKSIA 416

40

Query: 421 IITNDEMTEGLAQQTNRGVSKEIAEQAKHLDAQVLASDAFFPADNTERIAAAGIK 480
 I++ + MT+G-GAGQ NRVG+ KIAIECA + G+V+ SDAPP D +E A AGI
 Sbjct: 417 IVLADQQTIVGVGAQNNRVGAAKIAIEQAGSEKANGSVMSGDAFFMGDTVELAAKAGIT 476

45

Query: 481 AIIQPGGSVDRQESIDANNKHGLTMIPTGVRRPH 515
 AIIQPGGS+RD+ESI+ A+KHG+ M+PTGVRRPH+H
 Sbjct: 477 AIIQPGGSIRDEBSIENAKHGIDAMVPTGVRRPHK 511

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 959> which encodes the amino acid sequence <SEQ ID 960>. Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence

50

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2932 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

55

An alignment of the GAS and GBS proteins is shown below:

Identities = 500/515 (97%), Positives = 507/515 (98%)

60

Query: 1 MTKRALISVDSKGIIDFAKELKNLGNIIISTGGTKVALDDAGVETIADVDVTFPPE@D 60
 MTKRALISVDSKGIIDFAKELKNLGNIIISTGGTKV LDAGVETIADVDVTFPPE@D 60
 Sbjct: 1 MTKRALISVDSKGIIDFAKELKNLGNIIISTGGTKVLDAGVETIADVDVTFPPE@D 60

Query: 61 GRVKTLPNHNHOGLLARRDDSHLQAAXNNIELIDLVVNNLYPFKEITILRPDVTYDLAV 120

-381-

```

GRVVTLLHPNTHGGLARRDADSHLQAQKNNIELIDLVVVNLVYPFKSTILRPD+TYDLAV
Sbjct: 61 GRVVTLLHPNTHGGLARRDADSHLQAQKNNIELIDLVVVNLVYPFKSTILRPD+TYDLAV 120

Query: 121 ENIDIGGQPSMLRSAAKNHASVTVVVDSDADYATVLGELADASQTTFKTRQLAAKAFRHTA 180
5 ENIDIGGQPSMLRSAAKNHASVTVVVD ADYATVLGELADA QTTF+TRQLAAK FRHTA
Sbjct: 121 ENIDIGGQPSMLRSAAKNHASVTVVVDSDADYATVLGELADAGQTTFETRQLAAKAFRHTA 180

Query: 181 AYDALIAEYFTQVGEAKPEKLTITTYDLQAMRYGENPQQADPFYQKALPTDYSIASAQ 240
AYDALIAEYFT QVGEAKPEKLTITTYDLQAMRYGENPQQADPFYQKALPTDYSIASAQ
10 Sbjct: 181 AYDALIAEYFTQVGEAKPEKLTITTYDLQAMRYGENPQQADPFYQKALPTDYSIASAQ 240

Query: 241 LAGKELSFNNIRDADAARIIRDFPKDSPTVVALKHMNPGSIGQADDIETAWDYAEADPV 300
LAGKELSFNNIRDADAARIIRDFPKD PTVVALKHMNPGSIGQADDIETAWDY Y+ADPV
15 Sbjct: 241 LAGKELSFNNIRDADAARIIRDFPKDRPTVVALKHMNPGSIGQADDIETAWDYTYKADPV 300

Query: 301 SIFGGIIVLNRVDAATAKQHPITPLETIIAPSYSERAAILTNKKKNLRILELPDQA 360
SIFGGI+VLNRVDAATA+KQHPITPLETIIAPSYSERAAILTNKKKNLRILELPDQA
20 Sbjct: 301 SIFGGIIVLNRVDAATAKQHPITPLETIIAPSYSERAAILTNKKKNLRILELPDQA 360

Query: 361 ASEVEAEYTGVOGLLVQNCQVAENPSIDWCQVTRDQPTQBEATALEFAWKAIKYKXSG 420
ASEVEAEYTGVOGLLVQNCQVAENPSIDWCQVTRDQPTQBEATALEFAWKAIKYKXSG
25 Sbjct: 361 ASEVEAEYTGVOGLLVQNCQVAENPSIDWCQVTRDQPTQBEATALEFAWKAIKYKXSG 420

Query: 421 IIITNDEHMTLGLGAGQTNRVGSVKIAIEQAQKDLGAVLASDAFFPADNIEBIAAGIK 480
IIITNDEHMTLGLGAGQTNRVGSVKIAIEQAQKDLGAVLASDAFFPADNIEBIAAGIK
30 Sbjct: 421 IIITNDEHMTLGLGAGQTNRVGSVKIAIEQAQKDLGAVLASDAFFPADNIEBIAAGIK 480

Query: 481 AIIQPGSGSVRCQSIDAANKHGLTMIPTGVVRHFRH 515
AIIQPGSGSVRCQSIDAANKHGLTMIPTGVVRHFRH
30 Sbjct: 481 AIIQPGSGSVRCQSIDAANKHGLTMIPTGVVRHFRH 515

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 298

35 A DNA sequence (GBSx0326) was identified in *S. galactiae* <SEQ ID 961> which encodes the amino acid sequence <SEQ ID 962>. This protein is predicted to be similar to antibiotic resistance protein. Analysis of this protein sequence reveals the following:

```

Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1842 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
45 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:CAB12342 GB:Z99106 similar to antibiotic resistance protein
[Bacillus subtilis]
Identities = 65/263 (24%), Positives = 117/263 (43%), Gaps = 34/263 (12%)

Query: 5 KNLEIVESIFGD-WDETIINSQV-QGIMGEVFDLQPKSLAKLGRKSSFGFLAGQPT 62
K ++++F D + T ++S + Q I G V+ D PKS +G +S F+AG
Sbjct: 10 KKYSSSLKTMFDKCYPTFVYSILDQITPGVAYADQTFPKSFF--IGTESGIYFIADQDG 67

Query: 63 -----LFILEVCSGSDIILVQPKGWSDLIESTYQONAHSPKRYATYKDTLPERS 112
+ +V S + L W +++ + + +R A + +
Sbjct: 68 NRDFHDFIAGYIEQVSKSRFTLPSSSDITWDSVLKPKILKDLNQMRRAAFSY----QP 122

Query: 113 RLEKFPVTQLNPGFELRAIDKV-----YNSCLEKWSQDLVANYATYQYKKGIGIYVV 166
+ K QLP G L+ IDE + +NS +E+ + + +G G+ V
60 Sbjct: 123 KSPKKTQLPKGLVLRIDSDIISHSTAPNSAYEY-----WNSVQSPASKQGFPAV 175

```

-382-

Query: 167 YYQGNIIAGASSYSTYKNGIEIVDTHPOFRRRLGATIVAAQLIILPCLDKGIYPSWDAH- 225
 + ++++ +S N E+++ T ++R GLA VA + I C++ GI PSWD
 Sbjct: 176 LGHNVVSECTSIPLGHNRAEMDIITTEYRGLGLAYCANRFLAPCMENGIVPSWDCDI 235

Query: 226 -TRTSINLSKGLGYEPSHEYIAY 247
 +S+ L+ KLG++ EY Y
 Sbjct: 236 CNNSSTALAAKLGFKTVTEYTY 258

10 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 299

A DNA sequence (GBSx0328) was identified in *S.galactiae* <SEQ ID 963> which encodes the amino acid sequence <SEQ ID 964>. This protein is predicted to be phosphoribosylglycinamide formyltransferase homolog (purN). Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0736 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 965> which encodes the amino acid sequence <SEQ ID 966>. Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence

30 INTEGRAL Likelihood = -0.53 Transmembrane 75 - 91 (75 - 91)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1213 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CA04374 GB:AJ000883 purD [Lactococcus lactis]
 Identities = 236/419 (56%), Positives = 301/419 (71%), Gaps = 7/419 (1%)

40 Query: 50 LKLLVGGSGGRHIAIAKLLASGVDQVFPVAPNDGNTLDGLDVLNVIVSEHSRLIAFAK 109
 +K+LV+GSGGRHIA+AKK + S V+V+VAPGN GM DG+ +V+I + +L+ FA+
 Sbjct: 1 MKLLVGGSGGRHIAKAPMSQVIEVFVAPNGNMRDGIQIVHISLSDKLVKFAQ 60

45 Query: 110 ENESISWAFIGFDALAGAVDDFNSAGLRAPGPTKAAAELEMSKDFAKEINVKYNVPTAA 169
 I F+GP+ AL G+VD F A L RGP K AAKLE SKDFAK IM KY VPTA
 Sbjct: 61 NQMLIGLTVFGVEATLNGVVDFAFKARLPPIPGNMAAELESGKDFAKSIMKKYGVPTAD 120

Query: 170 YSTFSDPEKAKAYIESGQAPVIVKADGLALGKGVVVAETVBQAVSAACBMLDNKFPQDSG 229
 Y TF E A AY++E+G P+V+KADGLA GKGV VA +E A R ++ F S
 Sbjct: 121 YATFDSLEPALAYLDLRKGVPLVIKADGAAKGKGVTVAFDIETAKGALADI-----FSGSQ 175

Query: 230 ARVVIEEFLDGKEPFLFAFANGDKFYIMPTAQDHKRAFQDGKGNPTQGMGAYAFVHLPQ 289
 +VVIEEFLDGKEPFLF+P + K Y MP AQDHKRAF DGKGNPTQGMGAY+PV H+ +
 Sbjct: 176 GKVVIEEFLDGKEPFLFSFTHDGKIYMPAQDHKRAFQDGKGNPTQGMGAYSFVLHLSK 235

55 Query: 290 SVVDTAEMIVIRPVLGCMVAEGRPYLVGLVYGLILITADGPKVIEFNSRFGDPETQIILPR 349
 VV+ A+E +V+P + GM+ EG+ + GVLV GLILT DG K IEFN+RFGDPEIQ++LPR
 Sbjct: 236 EVVNEALEKVVKPTVAGMIEEGSKSFTGVLYAGLILITEDGVKTIERNARFGDPETQVVLPR 295

-383-

Query: 350 L T S D F A Q N I D D I M G S I S P Y I T W Q K D G V I L G V V V A S E G Y P P D Y E K G V L P E K T D G D I I T Y Y 409
 L S D A Q I D I + G E P + W + G V T I G V V V A + R G Y P + G + L F E + G + Y Y
 Sb j c t : 296 L E S D L A Q A I I D I L A G N E P T L E W L E S G V I L G V V V A R G Y P S Q A K L G L I L P E I P E G - L N V Y 354

Query: 410 A G V K P S N S E I L L S N G R V I M L V T T E D S V K A G Q D K I Y T Q L A Q D D T G L F T R N D I G S K A I 458
 A G V + E K + + L + S + G G R V Y + + T + V K + Q + Y + L + + G F Y R + D I G S + A I
 Sb j c t : 355 A G V S K N E N D Q - L I S S G G R V I L V S T G R D V K S T Q K L L Y E K L D K L E N D G F F T R E D I G S R A I 412

An alignment of the GAS and GBS proteins is shown below:

Identities = 172/182 (94%), Positives = 176/182 (96%)

Query: 1 M K I A V F A S G N S N F Q V I A E Q P Q V S F P S D H R D A Y V L E R A Q N L A I P S F A F E L K E F E N K A Y 60
 M K I A V F A S G N S N F Q V I A E Q F V S F P S D H R D A Y V L E R A Q N L A I P S F A F E L K E F E N K A Y
 Sb j c t : 1 M K I A V F A S G N S N F Q V I A E Q P P V S F P S D H R D A Y V L E R A Q N L A I P S F A F E L K E F E N K A Y 60

Query: 61 E Q A V D L L D K H E I D L V C L A G Y M K I V G E T L L A Y E G R I I N I H P T Y L P E F F G A I G K A D A E A 120
 E Q A + V D L L D K H E I D L V C L A G Y M K I V G E T L L A Y E R I I N I H P Y L P E F F G A I G I + D A E A
 Sb j c t : 61 E Q A I V D L L D K H E I D L V C L A G Y M K I V G E T L L A Y E R R I I N I H A Y L P E F F G A I G I E D A E A 120

Query: 121 G V D Q S G V T I H W D S G V D T G Q V I Q Q V H V P R L A D S L S F E T K I H E T Y Q L Y P A V L D S L G I K 180
 G V D Q S G V T I H W D S G V D T G Q V I Q Q V P R L A D S L S F E T K I H E T Y Q L Y P A V L D S L G + +
 Sb j c t : 121 G V D Q S G V T I H W D S G V D T G Q V I Q Q V K V P R L A D S L S F E T K I H E T Y Q L Y P A V L D S L G V E 180

Query: 181 R K 182
 R K
 Sb j c t : 181 R K 182

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 300

A DNA sequence (GBSx0329) was identified in *S. agalactiae* <SEQ ID 967> which encodes the amino acid sequence <SEQ ID 968>. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.59 Transmembrane 121 - 137 (121 - 137)

----- Final Results -----
 bacterial membrane --- Certainty=0.1235 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC16901 GB:AF016634 phosphoribosylformylglycinamide
 cyclo-ligase [Lactococcus lactis subsp. cremoris]
 Identities = 253/338 (74%), Positives = 286/338 (84%), Gaps = 4/338 (1%)

Query: 4 K N A Y A Q S G V D V R A G Y E V V R I K K H V A R T S R A G M S A L G S F G M F D L S Q T G V K E P V I L S G T 63
 + N A Y A + S G V D V R A G Y E V V R I K K H V A + T S R G V + G A L G S F G F D L S V R E P V I L S G T
 Sb j c t : 5 E N A Y A K S G V D V R A G Y E V V R I K K H V A K T E R L G V L G A L G S F G S F D L S V L D V K E P V I L S G T 64

Query: 64 D G V T K I M L A I K Y D K H D T I G Q D C V A M C V N D I I A A G A R P L Y F L D Y V A T G K E S P A N I E Q V V A 123
 D G V T K I M L A I + D K H D T I G D C V A M C V N D I I A A G A R P L Y F L D Y + A T G R N P K E Q V V A
 Sb j c t : 65 D G V T K I M L A I R A D K H D T I G D C V A M C V N D I I A A G A R P L Y F L D Y I A T G H I P E K I E Q V V A 124

Query: 124 G V A E G C I Q A S A L I G G E T A S M P G M Y G E D D Y D L A G F A V G V A E K S Q I I D G S K - V E S K D I L L G 182
 G V A E G C + Q A R A L I G G E T A S M P G M Y E D D Y D L A G F A V G V A E K S Q + I D G K V + G D + I L G
 Sb j c t : 125 G V A E G C I Q A R A L I G G E T A S M P G M Y D E D D Y D L A G F A V G V A E K S Q I I D G S K D V E G I L L G 184

Query: 183 L A S S G I H S K P S L V R R E A D Y F T G E V L P E L E C K Q K D V L L E P T R I Y V K A A L P L I K E E L V N 242
 L A S S G I H S K Y S L V R + V E A D + E L P E L + + L D I L F T + I Y V K L E L I K + +

-384-

Sbjct: 185 LABSGIHENGYSLVRKVFADFDNLSELPED-QSLIDTLLTPTKIYVKELLPLIKQNIK 243
 Query: 243 GIAHITGGGFIEIVPRMFADDLAAEIDEDKVPVLPFIFKALEKYGDIKHEMFRIPIFMGIG 302
 GIAHITGGGF IEN+PRMF + LAAEY R VLPIFKALRYG IKHEM+RIPIFMG+G
 Sbjct: 244 GIAHITGGGFHNLPRMFNGSLGARIVEGSDWLVLPFIFKALEKYGSIKHEMYRIPIFMGIG 303
 Query: 303 LMLDVNPNENVDVRKELLDEFPYIEGRIIIGKADSVVVK 340
 +++ V PEN ++K+L+ +EIG+++ + VVIK
 Sbjct: 304 MVIAVAPENAAALKELN--APEIGQNVNRQEAFFVVK 339

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 969> which encodes the amino acid sequence <SEQ ID 970>. Analysis of this protein sequence reveals the following:

Possible site: 38
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3236 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 321/340 (94%), Positives = 332/340 (97%)
 Query: 1 MSEKIIYAQSGVDVEAGVEVVERIKKHVARTERAGVMGALGGFGGMFDLSQTGVKEPVL 60
 MSEKIIYA+SGVDVEAGVEVVERIKKHVARTERAGVMGALGGFGGMFDLS+TGKVEPVL+
 Sbjct: 1 MSEKIIYA+SGVDVEAGVEVVERIKKHVARTERAGVMGALGGFGGMFDLSKTGVKEPVL 60
 Query: 61 SGTGVTGKMLAIKYDKHDTIGQDCVAMCVNDIIAAGAEPLFLDYVATGKNEPAKLEQ 120
 SGTGVTGKMLAIKYDKHDTIGQDCVAMCVNDIIAAGAEPLFLDY+ATGKN P KLE+
 Sbjct: 61 SGTGVTGKMLAIKYDKHDTIGQDCVAMCVNDIIAAGAEPLFLDYIATGKNPVLKEE 120
 Query: 121 VVAGVARGCVQASALIGETAENPMGYGDDYDLAGFAVGVAEKSQIIDGSRVKEGDIL 180
 VV+GVARGCVOA AALIGETAENPMGYG+DYDLAGFAVGVAEKSQIIDGSRVKEGDIL
 Sbjct: 121 VVSGVARGCVQGAALIGETAENPMGYGDDYDLAGFAVGVAEKSQIIDGSRVKEGDIL 180
 Query: 181 LGLASSGIHENGYSLVRRVFADYTGDEVLPLEGGKQKLDVLEPTRIYVKAALPLIKEEL 240
 LGLASSGIHENGYSLVRRVFADYTG E+LPLEGGKQKLDVLEPTRIYVKAALPLIKEEL
 Sbjct: 181 LGLASSGIHENGYSLVRRVFADYTGKELLPLEGGKQKLDVLEPTRIYVKAALPLIKEEL 240
 Query: 241 VNGIAHITGGGFIEIVPRMFADDLAAEIDEDKVPVLPFIFKALEKYGDIKHEMFRIPIFMG 300
 V GI HITGGGFIEI+PRMFADDLAAEIDEDKVPVLPFIFKALEKYGDIKHEMFRIPIFMG
 Sbjct: 241 VKIGIHITGGGFIEIPIPRMFADDLAAEIDEDKVPVLPFIFKALEKYGDIKHEMFRIPIFMG 300
 Query: 301 VGLMLDVNPNENVDVRKELLDEFPYIEGRIIIGKADSVVVK 340
 VGLML V+PNV+RVKELLDEFPYIEGRIIIGKAD SVVVK
 Sbjct: 301 VGLMLAVSPENVRKELLDEFPYIEGRIIIGKADSVVVK 340

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 301

A DNA sequence (GBSx0330) was identified in *S.galactiae* <SEQ ID 971> which encodes the amino acid sequence <SEQ ID 972>. This protein is predicted to be phosphoribosylpyrophosphate amidotransferase (purF). Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1112 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

-385-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AA12627 GB:U64311 phosphoribosylpyrophosphate amidotransferase
[Lactococcus lactis]
Identities = 340/470 (72%), Positives = 404/470 (85%), Gaps = 5/470 (1%)

Query: 3 YEVKSLNEBOGVFGIWGYPQAAQVITYFGLHSLQHRQOEGAGI+SNUNGKLYGYRNVLG 62
+E K+LNEBOGVFGIWGYPQAAQVITYFGLHSLQHRQOEGAGI+ N+NGKL +R +GL++
Sbjct: 37 FPAKTLNEBOGVFGIWGHPDAARLTYPGLHSLQHRQOEGAGILVNINGKLNHRGLGLVT 96

Query: 63 EVFKNQSELDNLITGMAAIGHVRYATAGSADIRNTQFFLYKFHDQGFALCHNGNLITNAIS 122
EVF+++ +L+ LTG++AIGHVRYATAGS+I NIQPF ++FHDG L HNKNLITNA S
Sbjct: 97 EVFRHKKDLEELTGSSAIGHVRYATAGSANINITQFPQFRFHDGSLGLAHNGNLITNAQSL 156

Query: 123 RKELEKQGAIFPNAASDTEILMHLIRSHNPSFMGKVKFALESTVKGGFAYLMTEDKLI 182
R ELEK GAIF+++SDTEILMHLIRSH+P FMG+VKEAL+TVKGGFAYL+MTE+ ++PA
Sbjct: 157 RCELEKSGAIFSSNSDTEILMHLIRSHHEPFGKVKFALESTVKGGFAYLMTENLIV 216

Query: 183 LDPNAPRPLSIGMNGAMVISSETCAFEVVGAKWVDVEGEVILIDDSGIGQIRYIDE 242
LDPN FRPLSIG+M NGA V++SETCAF+VUGA W++DV+PGE+I I+D GI D++TD
Sbjct: 217 LDPNGFRPLSIGMNGALVASETCAFVDVGATWQDVQGEITINDGSHVQDFTDS 276

Query: 243 TQLAICSMETVYFARPDSTIHGVNHTARKMGKRLAQEPKQADIVIGVFNLSLSAAMG 302
T + ICSMET+YFARPDI I GVNHTARK GK LAQE K DADIVIGVFNLSLSAA G
Sbjct: 277 TNNCTICSMETVYFARPDSTIAGNVNHTARKSGKILAEKADIVIGVFNLSLSAAG 336

Query: 303 FAEEBGLNEMGLVKNGYQRTFIQTQELREQGVMMKLSAVSGVVGKRVVMIDSEIVR 362
+AEEBGLP EMGL+KNQY RTFIQTQELREQGVMMKLSAV GVV+GRV+M+DSEIVR
Sbjct: 337 YAEESGLPNEGMLIKYQVARTFIQTQELREQGVMMKLSAVSGVVGKRVVMIDSEIVR 396

Query: 363 GTTSRRIVGLIREAGATEVHVVAISPELKYPCFYGIDIQTRRELISANHAHVDVCHIGA 422
GTTSRRIV LL++AGA EVHVVAISP LKYPCTFYGIDIQ R ELI+A H DE+ + IGA
Sbjct: 397 GTTSRRIVKLLDGAAGAEVHVVAISPAKYPCFYGIDIQDRDELIAHTHTDIREAIGA 456

Query: 423 DSLTYLSIDGLIKSIGLETKVPMGGLCVAYFDGRHYPTPLNDYEVTEVLSL 472
DSLTYLS GL+++IG + LC++YFDG YPTPLNDYEV +YL SL
Sbjct: 457 DSLTYLSQGLVEAIG-----HDKLCLSYFDGEVYPTPLNDYEVTEVLSL 500

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 973> which encodes the amino acid sequence <SEQ ID 974>. Analysis of this protein sequence reveals the following:

Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0610(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 473/484 (97%), Positives = 481/484 (98%)

Query: 1 MTEYKSLNEBOGVFGIWGYPQAAQVITYFGLHSLQHRQOEGAGI+SNUNGKLYGYRNVL 60
MTEYKSLNEBOGVFGIWGYPQAAQVITYFGLHSLQHRQOEGAGI+SNUNGKLYGYRNVL
Sbjct: 20 MTEYKSLNEBOGVFGIWHGPQAAQVITYFGLHSLQHRQOEGAGISNUNGKLYGYRNVL 79

Query: 61 LSEVFNQSELDNLITGMAAIGHVRYATAGSADIRNTQFFLYKFHDQGFALCHNGNLITNAI 120
LSEVFNQSELDNLITGMAAIGHVRYATAGSADIRNTQFFLYKFHDQGFALCHNGNLITNAI
Sbjct: 80 LSEVFNQSELDNLITGMAAIGHVRYATAGSADIRNTQFFLYKFHDQGFALCHNGNLITNAI 139

Query: 121 SRKELEKQGAIFPNAASDTEILMHLIRSHNPSFMGKVKFALESTVKGGFAYLMTEDKLI 180
R KELEKQGAIFPNAASDTEILMHLIRSHN PSFMGKVKFALESTVKGGFAYLMTEDKLI
Sbjct: 140 SRKELEKQGAIFPNAASDTEILMHLIRSHNPSFMGKVKFALESTVKGGFAYLMTEDKLI 199

-386-

Query: 181 AALDPNAFPLSIGQMGNAMVISSETCAFEVVGAKWVRD/EPGEVILIDGGIQCRRYT 240
 AALDPNAFPLSIGQMGNAMVISSETCAFEVVGAKWVRD/VKPEVILIDD GTQCRRYT
 Sbjct: 200 AALDPNAFPLSIGQMGNAMVISSETCAFEVVGAKWVRD/EPGEVILIDGGIQCRRYT 259

5 Query: 241 DETQLAICSMIEVYFARPDSTIHGVNHTARKNMGRKLAQEFKQADIVIGVNSSLSAA 300
 DETQLAICSMIEVYFARPDSTIHGVNHTARKNMGRKLAQEFKQADIVIGVNSSLSAA
 Sbjct: 260 DETQLAICSMIEVYFARPDSTIHGVNHTARKNMGRKLAQEFKQADIVIGVNSSLSAA 319

10 Query: 301 NGFARESGLPHEMLVGNQYTKRTFIQPTQLRECGVRMKLSAVSGVVGKRVVMIDDSI 360
 NGFARESGLPHEMLVGNQYTKRTFIQPTQLRECGVRMKLSAVSGVVGKRVVMIDDSI
 Sbjct: 320 NGFARESGLPHEMLVGNQYTKRTFIQPTQLRECGVRMKLSAVSGVVGKRVVMIDDSI 379

Query: 361 VRGITSRRIVGLLRAGATFEVHVAIASPELKYPCFYGIDICTRRELISANHAVDEVCDII 420
 VRGITSRRIVGLLRAGATFEVHVAIASPELKYPCFYGIDICTRRELISANHAVDEVCDII
 Sbjct: 380 VRGITSRRIVGLLRAGATFEVHVAIASPELKYPCFYGIDICTRRELISANHAVDEVCDII 439

15 Query: 421 GADSLTYLSIDGLISIGLETKAPNGGLCAVFDGHYPTPLVDYEEYLRSLSEKTSFYI 480
 GADSLTYLSIDGLISIGLETKAPNGGLCAVFDGHYPTPLVDYEEYLRSLSEKTSFYI
 Sbjct: 440 GADSLTYLSIDGLISIGLETKAPNGGLCAVFDGHYPTPLVDYEEYLRSLSEKTSFYI 499

20 Query: 481 QKVK 484
 QKVK
 Sbjct: 500 QKVK 503

- 25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 302

A DNA sequence (GBSx0331) was identified in *S.galactiae* <SEQ ID 975> which encodes the amino acid sequence <SEQ ID 976>. Analysis of this protein sequence reveals the following:

30 Possible site: 28
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 35 bacterial cytoplasm --- Certainty=0.4797 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 303

A DNA sequence (GBSx0332) was identified in *S.galactiae* <SEQ ID 977> which encodes the amino acid sequence <SEQ ID 978>. Analysis of this protein sequence reveals the following:

45 Possible site: 13
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 50 bacterial cytoplasm --- Certainty=0.3489 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 304

A DNA sequence (GBSx0333) was identified in *S. galactiae* <SEQ ID 979> which encodes the amino acid sequence <SEQ ID 980>. Analysis of this protein sequence reveals the following:

Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.1690 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

15 >GP:CA12194 GB:AL445066 phosphoribosylformylglycinamide synthase
related protein [Thermoplasma acidophilum]
Identities = 199/746 (26%), Positives = 329/746 (43%), Gaps = 103/746 (13%)

20 Query: 202 ADD--FAAYKAEGGLAMEVDLLFIQDYFKSIGRVPTELEKVLDTYMSHCRTTFETE 259
ADD A GLA+ +D++ ++ YF+ +GR P + E+ + MSAC + + +
Sbjct: 11 ADDARLKAISKRLGLALSLEMGAVRSYFTELGRDPIDRETHAQAQSWSEHCYSYKSKY 70

25 Query: 260 LGRIDFSASKFQQLQATYDYKIAMRDELGRSEKPTLMDMATIPQRYERANGRLDMEV 319
LK K+ L+ Y +AM D+ G
Sbjct: 71 LK-----KYLGLATDYT-ILAMEDDAG----- 92

30 Query: 320 SDEINACSVIEIVDVGVKEPMLLMFINETNHPTIEPPGGAATCIGGAIRDPLSGRSY 379
VD DG + + K E+NHNP+ +EP+GGAAT IGG +KD L +
Sbjct: 93 -----VVDPDG---EYAYVLIMESHNPISAVEPYGGAATGIGGIVDVLQWQAQ 138

35 Query: 380 VYQMRISGAGDITPIAETRAGLKPOQVISTKANGYSSYGNQIGLATTYVREYFHPGF 439
+ GD+++ E GL + I G YGN+IG+ YF +
Sbjct: 139 PVALIDSLFLGDVSSDRYR---GLLSPRYIPGGVVGIRDYGNRIGIPNVAGSLYFKLY 195

40 Query: 440 VAKRMELGAVVGAAPKENVREKP-EGADVVLLEGKTRGDVGAGATGSSKVTVESVET 498
+ + VG ++ +VR K + GDV+VL+GKTRGDG+ G+S + ++
Sbjct: 196 NSNPLNAGVCGVIRDRIVRSKSYKPGDVLVLMGKTRGDGIGVNFASITLG-KVTKS 254

45 Query: 499 AGAEVQKNAIEERKIQRLFRDGNVTRLLIKSNIDFGAGGVCAIGELAD---GLEILD 554
+ +Q GN I E+ + + N LI+ D G GG+ A E+ G E I LD
Sbjct: 255 SRLAIQLGNPIVEQPMIKVLEANDAGLIRAMIDGGGLSSAATMYVAGFGAEITLD 314

50 Query: 555 KVLKYQGLNGTELAISESQRMSVVPVGSVDVAFIAACNKENIDAVV+ATVTEKPLNM 614
+ LK ++G EI ISSESQRN + P DV+ K N+D V+ VT + +
Sbjct: 315 DIKLKSNMGSQWISSESQRNLMCECPYEDVKIRQLAEKNWLDPSVIGQVADRIRV 374

55 Query: 615 TWNGETIVDLERCFDLTGK-VVVDAKVVDKDLTFEARTISATLEADMIKVLSDLNH 673
+ I+D++ FLD + V + K V+K +TVE+ E L + + ++ LN
Sbjct: 375 YYKKRKIIDMDIRFLDGSPPYQPYRIKEVEKSVTEQ---EPEDLMSVPRDMARLNT 430

60 Query: 674 ASQKGLQTLFDSSVGRSTVNHPIGGR-YOITPTSSVQKLPVQYGVTTTASVMAQGYNFP 732
++ + +D +V St+ P GR + T +++V K P++ + V+ G P
Sbjct: 431 CARFNVVRQYDITVRGSTIVTFPVGRPNKTHADTVIK-PLNSM--RGLVITSGSRPN 487

Query: 733 IAEWSPHYKAAYAVIEATARIWATGADWSRARFSYQYFERMDQAEKFPQVQSALLGSI 792
+ PY G + EA ++TG R ++ E QQ V ++
Sbjct: 488 MVSVDVPTAGTILTAAYKNLISG---GRPHSVVDALNPNPKEKEIMQPFVESVRAIG 544

Query: 793 EAQIQGLPSIGGKDSMSQTFRELVPTPLVAPGVTTADS-RKVLSPFKAAGENY--- 848
+ + GLP + G S + + + PT V D R+ + K + G IY
Sbjct: 545 DFCRKAGLFPVAGNVSPYNEKRTDIMPETTIMVGLIDVRRSETTYMKSGNATYLIIG 604

-388-

Query: 649 -----YIPQAIISRDIDFDLIKMP--SQFRAIQAHKITAASAVKYGG 890
 Y G + D+D +F S+ + I + H +++ GG
 Sbjct: 605 EPCNLTGSEYSFMKGYTGGLFAPDLDELTRDFLSSKADMLLSSHDVSS-----GG 658

5 Query: 891 VLESIALMTGPNRIGASVEIAHDES 916
 + +L+ M+PG+ IG V+I+ ++
 Sbjct: 659 LPAALSEMSPGSGIGPHVDISNVSA 684

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 981> which encodes the amino acid
 10 sequence <SEQ ID 982>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 15 bacterial cytoplasm --- Certainty=0.1415 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

20 Identities = 1219/1256 (97%), Positives = 1226/1256 (97%)

Query: 11 SSYFVFAFLSDLVSYMKRIFVEKKADPGIKSASLVKELTHNLQLASLADLRIVQVYDVF 70
 SSYF VAPLSDLVSYMKRIFVEKKADPGIKSASLVKELTHNLQL SLK LRIVQVYDVF
 25 Sbjct: 2 SSYFFVAPLSDLVSYMKRIFVEKKADPGIKSASLVKELTHNLQLTSLKLRIVQVYDVF 61

Query: 71 NLAEDLLARAEEKHIFSSQVTDRLITAEITAEELDKVAFPAIEALGQPDFQRAASSQEALL 130
 NLAEDLLARAEEKHIFSSQVTD LITE EITAEELDKVAFPAIEALGQPDFQRAASSQEALL
 Sbjct: 62 NLAEDLLARAEEKHIFSSQVTDCLITETETAEELDKVAFPAIEALGQPDFQRAASSQEALL 121

30 Query: 131 LLSGDSQVKNVTAQLYLNVKDIABEALBAVICHYLLNFEVDSRFKDTITPLEVQAFPSVSDKT 190
 L GSDSQVKNVTAQLYLNVKDI EAEALBAVICHYLLNFEVDSRFKDTITPLE QAFPSVSDKT
 Sbjct: 122 LFGSDSQVKNVTAQLYLNVKDIABEALBAVICHYLLNFEVDSRFKDTITPLEQAFPSVSDKT 161

35 Query: 191 ISNLDFFETYQADDFAAYKAEQQLAMEVDOLLFTQYFKSIGRVRTETELKVLDTYMSDH 250
 I NLDFFETYQADDFA YKAEQQLAMEVDOLLFTQ+YFKSIG VPTETELKVLDTYMSDH
 Sbjct: 182 ISNLDFFETYQADDPATYKAEQQLAMEVDOLLFTQYFKSIGCVPTETELKVLDTYMSDH 241

Query: 251 CRHTTFETELKNIDPSASKFOKQLQATYDKYIAMDRELGRSEKPTLMDMATIRGYERA 310
 CRHTTFETELKNIDPSASKFOKQLQ TYDKYIAMDRELGRSEKPTLMDMATIRGYERA
 40 Sbjct: 242 CRHTTFETELKNIDPSASKFOKQLQATYDKYIAMDRELGRSEKPTLMDMATIRGYERA 301

Query: 311 NGRLLDMEVSDEINACSVIEIVDVGVKEPWLIMFQNETHNHPTIEBPGGAATCIGGAI 370
 NGRLLDMEVSDEINACSVIEIVDVGVKEPWLIMFQNETHNHPTIEBPGGAATCIGGAI
 45 Sbjct: 302 NGRLLDMEVSDEINACSVIEIVDVGVKEPWLIMFQNETHNHPTIEBPGGAATCIGGAI 361

Query: 371 RDLPLSGRSVYVQAMRISGADITTFIASTRACKLPQVVISKTAHGYSSYGNQIGLATTY 430
 RDLPLSGRSVYVQAMRISGADITTFIASTRACKLPQVVISKTAHGYSSYGNQIGLATTY
 Sbjct: 362 RDLPLSGRSVYVQAMRISGADITTFIASTRACKLPQVVISKTAHGYSSYGNQIGLATTY 421

50 Query: 431 VREYTHPGFVAKRMELGAVUGAAPKENVREKPEAGDVVVLLGGKTRDGVGATGSSKV 490
 VREYTHPGFVAKRMELGAVUGAAPKENVREKPEAGDVV+LGGKTRDGVGATGSSKV
 Sbjct: 422 VREYTHPGFVAKRMELGAVUGAAPKENVREKPEAGDVVLLGGKTRDGVGATGSSKV 481

55 Query: 491 QTVSVESTAGAEVOKGNAIEERKIQRIFRDGNVTRLIKKSNDFGAGGVCVAIGELADGLE 550
 QTVSVESTAGAEVOKGNAIEERKIQRIFRDGNVTRLIKKSNDFGAGGVCVAIGELADGLE
 Sbjct: 482 QTVSVESTAGAEVOKGNAIEERKIQRIFRDGNVTRLIKKSNDFGAGGVCVAIGELADGLE 541

Query: 551 IDLDKVPKYQGLNGTEIAISSQSRMSVVVPGSDVDAPIACNKENIDAVVVATVTEKP 610
 IDLDKVPKYQGLNGTEIAISSQSRMSVVV P+DVDAPACNKENIDAVVVATVTEKP
 60 Sbjct: 542 IDLDKVPKYQGLNGTEIAISSQSRMSVVVRPNDVDAPIACNKENIDAVVVATVTEKP 601

Query: 611 NLVMTWNGETIIVDLERFLDTNGVRVVDAKVVDKDLTVPEARTTSATLEADLKVLSD 670
 NLVMTWNGE IIVDLER FLDTNGVRVVDAKVVDKDLTVPEARTTSATLEAD LKVLSD
 Sbjct: 602 NLVMTWNGEIIIVDLERRFLDTNGVRVVDAKVVDKDLTVPEARTTSATLEADTLKVLSD 661

65

Query: 671 LNHASQKGLQTFIDSSVGRSTVNHPIGGRYQITPTRESSVKQLPVQYGVTTTASVMAQQYN 730
 LNHASQKGLQTFIDSSVGRSTVNHPIGGRYQITPTRESSVKQLPVQYGVTTTASVMAQQYN
 Sbjct: 662 LNHASQKGLQTFIDSSVGRSTVNHPIGGRYQITPTRESSVKQLPVQYGVTTTASVMAQQYN 721

5 Query: 731 PYIAEWSFYHGAAYAVIETARLVATGADWSRAPSQYQYFERMDKQAEPPQVQVSAIG 790
 PYIAEWSFYHGAAYAVIETARLVATGADWSRAPSQYQYFERMDKQAEPPQVQVSAIG
 Sbjct: 722 PYIAEWSFYHGAAYAVIETARLVATGADWSRAPSQYQYFERMDKQAEPPQVQVSAIG 781

10 Query: 791 SIERAQIQLGLPSIGGKDSMSGTFEELTVPPITLVAPGVTTADSRKVLSEPKKAMGENIYYI 850
 SIERAQIQLGLPSIGGKDSMSGTFEELTVPPITLVAPGVTTADSRKVLSEPKKAMGENIYYI
 Sbjct: 782 SIERAQIQLGLPSIGGKDSMSGTFEELTVPPITLVAPGVTTADSRKVLSEPKKAMGENIYYI 841

Query: 851 PGQAISEDIDFDLIXNPSQFPAIQAHKITAASAVKYGVLESALMTGNRIGASVEI 910
 PGQAISEDIDFDLIXNPSQFPAIQAHKITAASAVKYGVLESALMTGNRIGASVEI
 Sbjct: 842 PGQAISEDIDFDLIXNPSQFPAIQAHKITAASAVKYGVLESALMTGNRIGASVEI 901

15 Query: 911 AELDSLLTAQLGGFVFTSVEEIAADVVKIGQTQADFTTVKGNLQAGSLLSAFEGKLEEV 970
 AELDSLLTAQLGGFVFTSVEEIAADVVKIGQTQADFTTVKGNLQAGSLLSAFEGKLEEV
 Sbjct: 902 AELDSLLTAQLGGFVFTSVEEIAADVVKIGQTQADFTTVKGNLQAGSLLSAFEGKLEEV 961

20 Query: 971 YPTFEQVDAIEEVPVAVSDVVIKAKELIEKPVVYIPVFGTNSYDSAKAFEGVQASVN 1030
 YPTFEQVDAIEEVPVAVSDVVIKAKELIEKPVVYIPVFGTNSYDSAKAFEGVQASVN
 Sbjct: 962 YPTFEQVDAIEEVPVAVSDVVIKAKELIEKPVVYIPVFGTNSYDSAKAFEGVQASVN 1021

25 Query: 1031 LVPPVTLNEAAIAESVDITMVAIAKANIIFAGGPSAADEPDGSAKFTVNIILNEKVRRA 1090
 LVPPVTLNEAAIAESVDITMVAIAKANIIFAGGPSAADEPDGSAKFTVNIILNEKVRRA
 Sbjct: 1022 LVPPVTLNEAAIAESVDITMVAIAKANIIFAGGPSAADEPDGSAKFTVNIILNEKVRRA 1081

30 Query: 1091 IDSFIEKSGLIIGICNGFQALVKSGLLPYCNFERAGETSPTLFYNDANQHVAKMVEVTRIA 1150
 IDSFIEKSGLIIGICNGFQALVKSGLLPYCNFERAGETSPTLFYNDANQHVAKMVEVTRIA
 Sbjct: 1082 IDSFIEKSGLIIGICNGFQALVKSGLLPYCNFERAGETSPTLFYNDANQHVAKMVEVTRIA 1141

Query: 1151 NINSPWLAGVEVGDIHVIPVSHGSGKFVVSASEFAELRDNGQIWSQVDFDQGPSMDSKY 1210
 NINSPWLAGVEVGDIHVIPVSHGSGKFVVSASEFAELRDNGQIWSQVDFDQGPSMDSKY
 Sbjct: 1142 NINSPWLAGVEVGDIHVIPVSHGSGKFVVSASEFAELRDNGQIWSQVDFDQGPSMDSKY 1201

35 Query: 1211 NFGSVNATIGITSKNGQIIGKMGHSERWEDGLFNQIPGNKDQLFSAVKYFTGK 1266
 NFGSVNATIGITSKNGQIIGKMGHSERWEDGLFNQIPGNKDQLFSAVKYFTGK
 Sbjct: 1202 NFGSVNATIGITSKNGQIIGKMGHSERWEDGLFNQIPGNKDQLFSAVKYFTGK 1257

40

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 305

A DNA sequence (GBSx0334) was identified in *Sagalactiae* <SEQ ID 983> which encodes the amino acid sequence <SEQ ID 984>. This protein is predicted to be phosphoribosylaminoimidazole-succinocarboxamide synthase (purC). Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4783 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

55 The protein has homology with the following sequences in the GENPEPT database:

>P:AAA03540 GB:115190 SAICAR synthetase [Streptococcus pneumoniae]
 Identities = 183/231 (79%), Positives = 203/231 (87%)

Query: 1 MTNQLIYTGKARDIYTKDENVIRTVYKQNTMLNGARKETIDGKAIANNQISSLIPEKL 60
 M+ QLIY+GKARDIY+T+DEN+I + YKQAT NG +KE I GKG LNNQISS IPEKL
 60 Sbjct: 1 MSFQLIYSGKARDIYTTEDENLIISTYKQDATFNGVKGQIAGKGVNINQISSFIFEKL 60

Query: 61 NMAGVVTHTYIEQISKNEQLNKKVDII PLEVLENNVTAGSFSKRPGVEEGHVLSTPIVEFY 120
 N AGV TH++E++S EQLAKKV IIPLEVLENN TAGSFSKRPGV+RG LSTPIVEFY
 5 Sbjet: 61 NMAGVATHPVEKLEDTBQLNKKVKIIPLEVLENNYTAGSFSKRPGVDEGIALSTPIVEFY 120

Query: 121 YKNDLNDPFFINDEHVKFLGIIVNDEEIAVLKGETRINELLKDWFAQIGLNLIDFKLEFG 180
 YKD+L+DPFFINDEHVKFL I +D++IAYLK E R INELLK WFA+IGL LIDFKLEFG
 10 Sbjet: 121 YKNDLDDPFFINDEHVKFLGIADQCIAYLKGEARRINELLKWFVAETIGLNLIDFKLEFG 180

Query: 181 FDKDGKII LADEFSPNCRMLMDADGNHMDKDVFRDLGSLTDVYQVLEKL 231
 FDKDGKII LADEFSPNCRMLMDADGNHMDKDVFR L G LTVYV++V EKL
 10 Sbjet: 181 FDKDGKII LADEFSPNCRMLMDADGNHMDKDVFRDLGSLTDVYVYVLEKL 231

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 985> which encodes the amino acid
 15 sequence <SEQ ID 986>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 20 bacterial cytoplasm --- Certainty=0.3935(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 221/234 (94%), Positives = 228/234 (96%)

Query: 1 MTNQLIYTGKAKDIYSTKDENVIRTVYKDAQTNLNGARKETIDGKALANQISSLIPEKL 60
 +TNQLIY GAKADIYSTKDENVIRTVYKDAQTNLNGARKETIDGKALANQISSLIPEKL
 30 Sbjet: 11 VTNQLIYTGKAKDIYSTKDENVIRTVYKDAQTNLNGARKETIDGKALANQISSLIPEKL 70

Query: 61 NMAGVVTHTYIEQISKNEQLNKKVDII PLEVLENNVTAGSFSKRPGVEEGHVLSTPIVEFY 120
 N AGVVTHTYIEQISKNEQLNKKVDII PLEVLENNVTAGSFSKRPGVEEGHVLSTPIVEFY
 Sbjet: 71 NKAGVVTHTYIEQISKNEQLNKKVDII PLEVLENNVTAGSFSKRPGVEEGHVLSTPIVEFY 130

Query: 121 YKNDLNDPFFINDEHVKFLGIIVNDEEIAVLKGETRINELLKDWFAQIGLNLIDFKLEFG 180
 YKD+L+DPFFINDEHVKFLGIIVNDEEIAVLKGETR INELLK WFAQIGLNLIDFKLEFG
 35 Sbjet: 131 YKNDLDDPFFINDEHVKFLGIIVNDEEIAVLKGETRINELLKGFQIGLNLIDFKLEFG 190

Query: 181 FDKDGKII LADEFSPNCRMLMDADGNHMDKDVFRDLGSLTDVYQVLEKLIAL 234
 FD++G IILADEFSPNCRMLMD +GNHMDKDVFRDLG+LTDVYQVLEKLIAL
 40 Sbjet: 191 FDKDGKII LADEFSPNCRMLMDKGNHMDKDVFRDLGSLTDVYQVLEKLIAL 244

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

45 Example 306

A DNA sequence (GBSx0335) was identified in *S.galactiae* <SEQ ID 987> which encodes the amino acid
 sequence <SEQ ID 988>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 50 bacterial cytoplasm --- Certainty=0.2779(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9457> which encodes amino acid sequence <SEQ ID 9458>
 was also identified.

The protein has homology with the following sequences in the GENPEPT database:

-391-

>GP:AC35700 GB:AF041468 acyl carrier protein [Guillardia theta]
 Identities = 27/75 (36%), Positives = 52/75 (69%)

5 Query: 12 MSRDEVFPEHLELLAQQLGDFQDLITPSSSLHDDLAIDSTALTEFTINLRDVFHLEIPDE 71
 N+ E+FEK+ ++ +CLEG + +T+++ +DL DG+ E ++ +E+ F+EIFD+
 Sbjct: 1 MNEQEIFEKVTIISBQLGVDRSQVTKDMFANDLGADSLDVELVMAIEEAPHIEFD 60

Query: 72 AVEHMSVQQLLDYI 86
 A E +S++QQ +D+I
 10 Sbjct: 61 AAEQISNLQQAVDFI 75

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 989> which encodes the amino acid sequence <SEQ ID 990>. Analysis of this protein sequence reveals the following:

Possible site: 24
 15 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1917(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 20 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 36/77 (46%), Positives = 57/77 (73%)

25 Query: 12 MSRDEVFPEHLELLAQQLGDFQDLITPSSSLHDDLAIDSTALTEFTINLRDVFHLEIPDE 71
 M+R E+FE+++ L+++Q + IT ++ L +DL+DSI L EFTIN+ED FH+ IDE
 Sbjct: 1 MTRQEIFERLINLIQKQSYLSVAITEQTWLNKDLAVDSIELVEFTINVEDEPHIAIPDE 60

Query: 72 AVEHMSVQQLLDYIIE 88
 VE M ++ +LDY+++
 30 Sbjct: 61 DVEDMVMFRDILLYLVQ 77

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

35 Example 307

A DNA sequence (GBSx0336) was identified in *S.galactiae* <SEQ ID 991> which encodes the amino acid sequence <SEQ ID 992>. This protein is predicted to be fatty acid/phospholipid synthesis protein (plsX). Analysis of this protein sequence reveals the following:

Possible site: 21
 40 >>> Seems to have no N-terminal signal sequence

INTERGRAL Likelihood = -0.64 Transmembrane 101 - 117 (101 - 117)

----- Final Results -----
 bacterial membrane --- Certainty=0.1256(Affirmative) < succ>
 45 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9455> which encodes amino acid sequence <SEQ ID 9456> was also identified.

50 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB13462 GB:Z99112 alternate gene name: ylpd [Bacillus subtilis]
 Identities = 174/329 (52%), Positives = 238/329 (71%), Gaps = 2/329 (0%)

Query: 8 KIAIDAMGGDIYAPKAIVBGNQASIDFSDIRVLYGDKKIEKYLTVT-RRVSIHTEEK 66
 +IA+DAMGGD+APKA+++GV + I F D+ + L GD+ IE +LT T +R+++H +E
 55 Sbjct: 2 RIAVDAMGGDIHAPKAVIDGVKIGRAFDULHITLWGDRTTIESHLTTSDRITVLHADEV 61

-392-

Query: 67 INSDEPAKAVRRKKQSSMVLGAKAVKDGVAQAFISAGNTGALLAAGLFPVVGRIKGVDRP 126
 I DEP +AVRRKK SSMVL A+ V + A A ISAGNTGAL+ AGLF+VGRIGK+DRP
 Sbjct: 62 IETDEPVRVAVRRKKSSMVLMAQEAENRADACTSAGNTGALMTAGLFTVGRIGKIDRP 121

Query: 127 GLMSTMTPLDGVGFMMLDGLANAENTASHLHQYAILGSPYAKNVRGIEVPRVGLLNNGTE 186
 L T+PT+ G GF +LD+GN+ HL QYAI+GS Y++ VRG+ PRVGLLN GTE
 Sbjct: 122 ALAAPTFLFVSGDGLLLDVGANVDKPEHLVQYALMGSVYSQQVRGVTSRVLGLNVGTE 181

Query: 187 EETKGDLSLHKEAYELLAAEPISFNIGNIEARDLMSVADVVVDGFTGNVLTKEBTGMS 246
 + RG+ L K+ +++L +INFIGN+EARL+ VADVVDGFTGN LKT+EG+A+S
 Sbjct: 182 DKKGHSLTKQTFQLIKETAMINFIGNVEARDLDDVADVVDGFTGNVLTKEBTGMS 241

Query: 247 IMGSLKSSIKSGGVKAKLGALLKDSLYOLKDSMDYSSAGGAVLPLGKAPVTKCHGSSDS 306
 I ++ + + + +K L A +L K L ++K M+Y+G GGA LPLGLAP++K HOSSDS
 Sbjct: 242 IFFOMR-DVMTSTLTSKLAALVLPKPKLRMKNMRYSNYGASLPLGLKAPVKAHSSDS 300

Query: 307 KAVYSTLKQVTRMLETOVVDQLVDAFTE 335
 AV+ ++Q R M+ V + + +E
 Sbjct: 301 NAVFAIRQAREMVSQNVALLIQEEVKEE 329

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 993> which encodes the amino acid sequence <SEQ ID 994>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.07 Transmembrane 121 - 137 (120 - 138)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1829 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9127> which encodes the amino acid sequence <SEQ ID 9128>. Analysis of this protein sequence reveals the following:

Possible cleavage site: 16
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.07 Transmembrane 95 - 111 (94 - 112)
 ----- Final Results -----
 bacterial membrane --- Certainty= 0.183 (Affirmative) < succ>
 bacterial outside --- Certainty= 0.000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty= 0.000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 254/330 (76%), Positives = 290/330 (86%)

Query: 6 MKKIAIDAMGGDYAPKAIVEGVNQALSDPSEIEVQLYQDQKIEKYLTVTERVSIITHEE 65
 NK+IAIDAMGGD APKAIVEGVNQAL PSDEI+QLYQDQ KI YL ++RV+IITHEE
 Sbjct: 27 MKKIAIDAMGGDNAPKAIVEGVNQALSDPSEIEIQLYQDQKINSYEDRVAIITHEE 66

Query: 66 KINSDEPAKAVRRKKQSSMVLGAKAVKDGVAQAFISAGNTGALLAAGLFPVVGRIKGVDR 125
 KI SDDEPAKAVRRKK++SMVLG AKAVK+G A A ISAGNTGALLA GLFPVVGRIKGVDR
 Sbjct: 87 KINSDEPAKAVRRKKQSSMVLGAKAVKDGKADAITISAGNTGALLAAGLFPVVGRIKGVDR 146

Query: 126 PGLMSTMTPLDGVGFMMLDGLANAENTASHLHQYAILGSPYAKNVRGIEVPRVGLLNNGT 185
 PGL+ST+PT+ G+GFMMLDGLANAENTA HLHQYAILGSPYAKNVRG I PRVGLLNNGT
 Sbjct: 147 PGLSTPTPTVLGLGFMMLDGLANAENTAKHLHQYAILGSPYAKNVRGIANPRVGLLNNGT 206

Query: 186 EETKGDLSLHKEAYELLAAEPISFNIGNIEARDLMSVADVVVDGFTGNVLTKEBTGMS 245
 EETKGD L K YELL A+ +I+P+GN+EAR+LMS VADV+V+DGF+GNVLTKE+BTG+
 Sbjct: 207 EETKGDRLKATYELLTADNTISFPVGNVEARELMSGVADVIVSDGFTGNVLTKEBTGMS 266

Query: 246 SINGSLKSSIKSGGVKAKLGALLKDSLYOLKDSMDYSSAGGAVLPLGKAPVTKCHGSSD 305

-393-

SIMG LK I SGG+K K+GA LLK SLY+K ++DYSSAGGAVLFGLKAP+VK HGSSD
 Sbjct: 267 SIMGQLKQIINSGETTKYIGASLLKSSLYEMKKTLDYSSAGGAVLFGLKAPVVRHGSSD 326
 Query: 306 SKAVISTLKQVRITMLETQVVDOLVDAFTDE 335
 KA++ST+KQVRITMLE+T VV QLV+ F E
 Sbjct: 327 VKATPSTIKQVRITMLDTNVVQGLVBEFAKE 356

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 308

A DNA sequence (GBSx0337) was identified in *S. agalactiae* <SEQ ID 995> which encodes the amino acid sequence <SEQ ID 996>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4668(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25 Example 309

A DNA sequence (GBSx0338) was identified in *S. agalactiae* <SEQ ID 997> which encodes the amino acid sequence <SEQ ID 998>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood =-12.84 Transmembrane 61 - 77 (55 - 82)
 INTEGRAL Likelihood =-10.14 Transmembrane 26 - 42 (19 - 51)
 INTEGRAL Likelihood =-9.77 Transmembrane 192 - 208 (186 - 211)
 INTEGRAL Likelihood =-5.79 Transmembrane 267 - 283 (262 - 286)
 INTEGRAL Likelihood =-3.77 Transmembrane 100 - 116 (99 - 116)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.6137(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9453> which encodes amino acid sequence <SEQ ID 9454> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA22372 GB:AL034446 putative transmembrane protein
 [Streptomyces coelicolor A3(2)]
 Identities = 47/154 (30%), Positives = 69/154 (44%), Gaps = 12/154 (7%)
 Query: 120 SGFVEISSNSFSFGPFFFLFLAYFIQSILTEILFRGYWITVTYFKGSFAGVLNCSMLF 179
 SG+ E+ S F+A + TRE++FRG + + + G++ + ++F
 Sbjct: 118 SGYYEVDGLSGVQGAIGLVGMA--AAANTEVVRFGVLFRITREHIGTYLALGILGLVF 175
 Query: 180 SPFHFRN----YGITATLALFLLGLTYPSILFNMKNTLFTVGHVHTWNPMTGCVLGN 234
 +H N +G AIA+ F+L ++ T+N+ GVH WNP G V

-394-

Subjct: 176 GIMHLINRDATAWGAALATATEAGFMIAAYAA---TRNLKLTIGVHFGNNFAAGGVFST 231

Query: 235 KVSGGSDSPUSLFRITNSFFALWNGGDFGREGGV 268

VSG L T S L GDFG BG V

5 Subjct: 232 VVSGNGDSEGLLDAT-MSGPKLLTGDFGREGSV 264

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 310

A DNA sequence (GBSx0339) was identified in *S.galactiae* <SEQ ID 999> which encodes the amino acid sequence <SEQ ID 1000>. Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have no N-terminal signal sequence

15

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2665(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20

A related GBS nucleic acid sequence <SEQ ID 9451> which encodes amino acid sequence <SEQ ID 9452> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

25

>GP:BA05088 GB:AP001511 unknown conserved protein [Bacillus halodurans]
Identities = 81/242 (33%), Positives = 124/242 (50%), Gaps = 3/242 (1%)

Query: 8 GLVLNNRYREDDKLKIFTEBGRMPFVKHAS--KSKFNAVLQPLTIAHFLKINDKG 65

G+V+ +Y E+K+V+FT GK +A ES+ AV Q T + + N G

Subjct: 7 GIVIKTVDYGESNKIVTVFTREYKIALMARGAKRPKSRITAVTQLFTYQMMFQKQA-G 65

30

Query: 66 LSYIDYKRVLAFOETNSDLPKLSYASYITSLADVAISDNVADALFLFKKTLIELIDG 125

L + + + +F+E +DLF+ SY SY+T L + D + LF L +T+ + +G

Subjct: 66 LGTLTGRIIQSFREVRMDLFRASYSYVITLTKLTKTEDEKRNPLFELLYQTITHYNEG 125

35

Query: 126 LDVEILLTNIFEVQLLERPGVALNFHDCVFCRVLGPPDFSHKYSGLLCPNHYKDERRNH 185

+D ++LT IFEV++ G+ CV C +P FS K +G LC KD

Subjct: 126 MDPFDVTRITFEVKMPTVAG:KPRLDQCVRSTVDPVPGFSIKEAGFLCKRCIEKDPHAYK 185

40

Query: 186 LDPNMLYLNNRFOSIQFDOLQTISVKPQMKLKIRQFLDMYDEYVGIHLKSKFIDDLSSWG 247

+ + L+ F L TIS+KPE K ++ + YDEV G+HLKS++F+D L S G

Subjct: 186 ITAQVAKLRLFLYHFDLRIGLTISLKPETKATLKTIIHQYDEYSGHLKSRFLDQLSSWG 247

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1001> which encodes the amino acid sequence <SEQ ID 1002>. Analysis of this protein sequence reveals the following:

45

Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1566(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50

An alignment of the GAS and GBS proteins is shown below:

Identities = 159/251 (63%), Positives = 210/251 (83%)

55

Query: 1 MRVSGTGLVLYNNRYREDDKLKIFTEBGRMPFVKHASKSKFNAVLQPLTIAHFLK 60

-395-

- M++++ G+VL+NRNYKEDDKLVKIFTE GK+MPFVKH S+SK ++++QPLTIA FI K
 Sbjct: 1 MQLTESLGLVLFNRYKEDDKLVKIFTEVAGKMPFVKHISRSKMSLIQPLTLDATFIK 60
- Query: 61 INDNGLSYIDYKEVIAPOETNSDLFLSYASYITISLADVAISDNVADAQLFIFLAKTLE 120
 +ND GLSY+ DY V ++ N+D+P+L+YASY+ +LAD AI+IN +D+ LF FLKKTIL+
 Sbjct: 61 LNDTGLSYVVDYSNVNTYTRYINNDIFRLAYASYVLALADAAADNDSSEHLFTFLKKTLD 120
- Query: 121 LIEDGLGYEILLNIFEVOLLERFGVALNFHDVCVCHRVGLPDPSPSKYSGLLCPNHYHID 180
 L+E+GLDYELLNIFP+QHL+RFQ++LNFR+C CHR LP DFRH++S +LC HYHD
 Sbjct: 121 LMEBGLOYEILLNIFPQILDRFGISLNFRCAICHTDLPDPSRFAVLCEHYHND 180
- Query: 181 ERNHLDPNMLYLINFQSIQFDLLQT+SVKPEMKLKIRQFLDMYDEYVGIHLKSKKFI 240
 ERNHLDPN++YL++RFQ I FDDL+TIS+ ++K K+RQF+D +Y +YVGI LKSK FI
 Sbjct: 181 NRRNHLDPNVTYLLSKFQKIT+FDLRTI SLNKDINKKLKQFIDELHDYVGIKLSKTFI 240
- Query: 241 DDLSSGGSIMK 251
 D+L WG IMK
 Sbjct: 241 DNLVNGDLMK 251
- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 311

- A DNA sequence (GBSx0340) was identified in *S. agalactiae* <SEQ ID 1003> which encodes the amino acid sequence <SEQ ID 1004>. This protein is predicted to be aromatic amino acid aminotransferase (patA). Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -3.13 Transmembrane 141 - 157 (140 - 159)

- Final Results -----
 bacterial membrane --- Certainty=0.2253 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- A related GBS nucleic acid sequence <SEQ ID 9449> which encodes amino acid sequence <SEQ ID 9450> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

- >GP:AAF06954 GB:AF146529 aromatic amino acid aminotransferase
 [Lactococcus lactis subsp. cremoris]
 Identities = 261/391 (66%), Positives = 323/391 (81%)
- Query: 38 MTLKGRFNKYLDRIEVSILROFDQSIDIPGMVKLTILGRPDFITPDHVEKAASDAIDNQ 97
 M L K+FN LD+IE+SLIROFDQ +S IP ++KLTLGEPDF TP+HVKA +AI+ NO
 Sbjct: 1 MDLLKGFNPLDKTEISILROFDQVSSPDIIKLTILGEPDFITPDHVGAGIAIENQ 60
- Query: 98 SYTTGNSGLLALRQWADFAKDKYNILNTPDCEILVTIGATEALSASLIALEAGDVLL 157
 S+YTG+GLL LRQA++F KY L+Y + EILVT+G TEA+S L++IL AGD VL+
 Sbjct: 61 SHYTMAGLELRLQWASEFLKKYGLSYAEDIELVTGVTEAISSVLSILVAGDEVLI 120
- Query: 158 PAFAYPGYEPVNLVAGADIVELDTRENDPRLTPEMLTALIOQCEKLVANLYNPINTG 217
 PAFAYPGYEP++ L G +VELDTR NDF LTPML+ AII++ K+KAV+INYP NPTG
 Sbjct: 121 PAFAYPGYEP LITLAGGSLVELDTRANDFVLTPEMLDQALIEREGKVKAVILNYPANPTG 180
- Query: 218 ITYSRQIEALAEVLKKYDIFVISEDEVSELYTTQQRVSIAYELENQTILNGLSKSEA 277
 +TY+R++I LARVLKK++FV+DRVSELY YT Q HVSIAYE P QTI++NGLSKSEA
 Sbjct: 181 VTYNREQIKDLARVLKKIEVFVIADEVSELYNTDQHVSIAYEAPETQILNGLSKSEA 240
- Query: 278 NTGWRVGLWYAEPAFIAQIKSHQYVIAASTISQFAGVEALSVGKNDTLPKQGYIKR 337
 NTGWR+GL++A +AQIIK+HOY+VT+AST SQFA +RAL G +D LPM++ Y+KRR

-396-

Subjet: 241 MTGWRIGLIPARELVAQIIKTHQYLVTSASTQSQFAITALKNGADDALPMKKEYLKR 300

Query: 338 DYIIDKMSKLGFKIIKPSGAFYIFAKIPDSYQDSFKPQDFAYQQAVALIPGVAPGKYG 397
DYII+KMS LGFKII+P GAFYIFAKIP QDSFKP DFA + AVALIPG+APG+YG

Subjet: 301 DYIIEKMSLGFIIKIPDGAFYIFAKIPADLSQDSFKPFAVDFAKSNAVALIPGIAFGQYG 360

Query: 398 EGYIRLSYAASNEVIETAMARLKVPMBSYBG 428

EG++RLSYAASN+VIE AMARL ++ G

Subjet: 361 EGFVRLSYAASNDVIEQAMARLIDYVTKKGG 391

There is also homology to SEQ ID 1006.

SEQ ID 1004 (GBS332) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 60 (lane 3; MW 50.7kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 67 (lane 4; MW 76kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 312

A DNA sequence (GBSx0341) was identified in *S. agalactiae* <SEQ ID 1007> which encodes the amino acid sequence <SEQ ID 1008>. This protein is predicted to be ribose-phosphate pyrophosphokinase (prsA).

Analysis of this protein sequence reveals the following:

Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3118 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9447> which encodes amino acid sequence <SEQ ID 9448> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA62181 GB:M92842 prs [Listeria monocytogenes]
Identities = 209/312 (66%), Positives = 266/312 (84%), Gaps = 3/312 (0%)

Query: 10 LKLFALSSNKELAKKVSQTIGIPLGQSTVRFQSDRIQVNIRESIRGHVFIQSTSSPV 69
LK+L+SN+ELA++++ +GI LG+S+V FSDRIQ+NIRESIRG HV+++QSTS+PV

Subjet: 10 LKIFSLNKNRELAETAKGVIELGNSVTFHSDGTQINTRESIRGCHVYVIQSTNTV 69

Query: 70 NDNLEILIMVDALKRASBSVSVVMPYGYARQQRKARSREPIT+KLVAN+LEVAGVDR 129
N NME+LIN+DALKASA +++VMPYGYARQQRKARSREPIT+KLVAN++E AG R

Subjet: 70 NQNLNELLIMVDALKRASATINIVMPYGYARQQRKARSREPITAKLVANLIE+TAGATR 129

Query: 130 LLTVDLHAAQIQGFDPIDHLMGAPLIADYDFRQGLVGDVVVVSPIHGVTRARKLAQ 189
++T+D+HA QIQGFDPIDHIL L++DYF + L GUD+VVVSPIHGGVTRARK+A

Subjet: 130 MTLDMHAPQIQGFDPIDHILNANRLSDYFSEHIL-GDULVVVSPIHGGVTRARKAD 188

Query: 190 CISTPIAIIKKRRSVTKMTSRVMNIIGMIGKKCILIDMDITAGTICHAADALAEAGA 249
LK PIATIIKKR + N +EVMNI+GN++GK CI+IDD+IDTAGTI AA AL EAGA

Subjet: 189 RLKAPIAIIKKR-- -PRNVARVMNIVGNVBEKVCIIIDDIIDTAGTITLAALKALRENGA 246

Query: 250 TAVVYASCTHPVLSGPALNTQNSAIEKLIVLDTIYLPFERLIDKIEQISLARLGEAIR 309
T VYA C+HPVLSGPA+ I+ S IEKL+V ++I LFE+ IDK+EQ+S+A L+GEAI+R

Subjet: 247 TKVYACCHPVLSGPMKRIEBSPTIEGLVNTSIALPEEKWIDKMEQLSVALLGEAIR 306

Query: 310 IHEKRLPLPLFE 321
+HE +S LFE

-397-

Sbjct: 307 VHENASVSSILF3 318

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1009> which encodes the amino acid sequence <SEQ ID 1010>. Analysis of this protein sequence reveals the following:

5 Possible site: 22
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2685 (Affirmative) < succ>
 10 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 298/322 (92%), Positives = 311/322 (96%)

15 Query: 1 MEEIMSYSNLKLFPALSSNKEIAKKVSTGTIGPLQSTVRQFSDGEIQVNIIESIRGHVVF 60
 +EE MSYB+LKLFPALSSNKEIA+KV+ +GI LG+STVRQFSDGEIQVNIIESIRGHVVF
 Sbjct: 1 LEEIMSYSLKLFPALSSNKEIAEKVASAMGICLGKSTVRQFSDGEIQVNIIESIRGHVVF 60

20 Query: 61 ILQSTSSPVNDNLMEILIMVDALKFASASVSVVMPFYGVARQDRKARSREPITSKLIVAN 120
 ILQSTSSPVNDNLMEILIMVDALKFASAE +SVVMPFYGVARQDRKARSREPITSKLIVAN
 Sbjct: 61 ILQSTSSPVNDNLMEILIMVDALKFASAEKISVVMPFYGVARQDRKARSREPITSKLIVAN 120

25 Query: 121 MLEVAGVDRLLTVDLHAAQIQGFDFIPVDHLGAPLIADYFTRQGLVGDDVVVSPDH3G 180
 MLEVAGVDRLLTVDLHAAQIQGFDFIPVDHLGAPLIADYFTR GLVG+DVVVSPDH3G
 Sbjct: 121 MLEVAGVDRLLTVDLHAAQIQGFDFIPVDHLGAPLIADYFTRHGLVGDDVVVSPDH3G 180

Query: 181 VTRARKLAQCLTPIAIIIDKRBSVTKONTSEVNTIGNKKKCLILDDWIDTAGTICHA 240
 VTRARKLAQ L+TPIAIIIDKRBSV TKONTSEVNTIG+ GKCKILIDWIDTAGTICHA
 30 Sbjct: 181 VTRARKLAQLTQPIAIIIDKRBSVDKONTSEVNTIGN+SGKKCLILDDWIDTAGTICHA 240

Query: 241 ADALAEAGATAVYASCTHPVLSPGPALENIQSAIEKLIVLDTTYLPKRLIDKIEQISIA 300
 ADALAEAGATAVYASCTHPVLSPGPALENIQ SAIEKLIVLDTTYLP+ERLIDKIEQISIA
 35 Sbjct: 241 ADALAEAGATAVYASCTHPVLSPGPALENIQSAIEKLIVLDTTYLPKRLIDKIEQISIA 300

Query: 301 ELIGEAIIRIHEKRPLSPLEFM 322
 +L+ EAIIRIHEKRPLSPLEFM
 Sbjct: 301 DLVAEAIIRIHEKRPLSPLEFM 322

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 313

A DNA sequence (GBSx0342) was identified in *S.agalactiae* <SEQ ID 1011> which encodes the amino acid sequence <SEQ ID 1012>. This protein is predicted to be a secreted protein. Analysis of this protein
 45 sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 50 bacterial cytoplasm --- Certainty=0.3751 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9277> which encodes amino acid sequence <SEQ ID 9278>
 55 was also identified.

The protein has homology with the following sequences in the GENPEPT database:

-398-

>GF:AAD00288 GB:U78607 putative secreted protein [Streptococcus mutans]
Identities = 111/157 (70%), Positives = 130/157 (82%), Gaps = 1/157 (0%)

Query: 1 MTAIGKQGVAGLESGQSELEAGNQLGELVBAQVQGLQITGLSTKIVARNSESLKQVRSAGK 60
+ I+QGV AL+QQ+AL+H+H+LEBA S LQ+ITGLSTKIVARNSESLK+QV RSAGK
Sbjct: 55 LITTCQGVSLQTQAGALQANRNLGELVBAQVATLQOQITGLSTKIVARNSESLKQVRSAGK 114

Query: 61 NL-TNYITNTINSLKVSDAHVNRVVALEHVSSANEKMLAQSRADKAALAEQIENQNAINT 119
N+TYIN I+NSKVSDA+HNV ALHVVSANEKMLQQR DKAA+QK ENQ AINT
Sbjct: 115 NRTSYTNTINSLKVSDAHVNRVVALEHVSSANEKMLQQRDKAAVSKQQRQENQNAINT 174

Query: 120 VAANKATEINPAALATQALAEPAQLGELVBAQVQGLQITGLSTKIVARNSESLKQVRSAGK 156
VAAN+I N AL TL+Q+ALAEPAQL L+AVTIT
Sbjct: 175 VAANKATEINPAALATQALAEPAQLGELVBAQVQGLQITGLSTKIVARNSESLKQVRSAGK 211

There is also homology to SEO ID 1014.

A related GBS gene <SEQ ID 8543> and protein <SEQ ID 8544> were also identified. Analysis of this protein sequence reveals the following:

```

20 Lipop: Possible site: -1  Crend: 3
    MoG: Discrim Score:      8.29
    GvH: Signal Score (-7.5): 0.8
        Possible site: 49
    >>> Seems to have a cleavable N-term signal seq.
    ALOM program  count: 0 value: 6.74 threshold: 0.0
25 PERIPHERAL Likelihood = 6.74 400
    modified ALOM score: -1.85

    *** Reasoning Step: 3

30 ----- Final Results -----
        bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000 (Not Clear)

```

35 The protein has homology with the following sequences in the databases:

32.8/56.3% over 439aa

Lactococcus lactis

[illegible]

-399-

```

      804      834      864      894      924      954      984      1014
EAKQIENQVNTVANKQAIENNKAALATQRAQLSAAQLLSAQLTTVQNEKASLIQAKQAQESAARIAAQAQAAAEAK
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
5  SQKSETVKRGTYNQVSLSSLSQSAQLTSQAQLKATVATNYQATITATQDKKQALLDEKKAANQAQESAIAIKQAAVYEAQ
      170      180      190      200      210      220      230      240

      1044      1065      1095      1125      1155      1185      1215
AQRAKAQAQESVA---KAQAQAQVESATAPTEVQTPRTIKPSNL/TATSSATTVAITTTATINPEKVTQPSVTKA--
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
10 QKEAAQAQAATAATAKAVEATSTSSASASSQAQFQVSTSDMTTSSASASSSSSSSSSSSSSSSSSSSSSSSSSSSSSAGG
      250      260      270      280      290      300      310      320

      1266      1296      1326      1347      1374      1401      1455
-VEAPKAVVSESTPRAVSKPVVRSYDSSNFTYMQOCT---WGA-KEMASVGVNTW-GNNMQAGASARAAG--YVGTITPRV
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
15 NTNSGTSTGNTGTTTGGSSGINSPIGNFYAGGGCITYVWQFQAAGGIYIRNIMPNGGQWASNGPAQGVLVVVGAAQGV
      330      340      350      360      370      380      390      400

      1503      1533      1563      1593      1623      1653      1683
QAVAWP----YDGGYGHVAVVTSVANHSSIQVMESHYAGNMSIGNYHSGSNFSAAGSVYIYIPN**ILRRSPVVSFLF
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
20 IASSFSADFVGYANSFVGHVAIVKSVNSDGTITIKBGGYGTITWGHRTVSASGVTFILMPN
      410      420      430      440      450      460

```

SEQ ID 8544 (GBS65) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 5 (lane 6; MW 47.5kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 13 (lane 3; MW 72kDa) and in Figure 175 (lane 2 & 3; MW 72kDa).

The GBS65-GST fusion product was purified (Figure 102A; see also Figure 191, lane 4) and used to immunise mice (lane 1 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 102B), FACS, and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 314

A DNA sequence (GBSx0343) was identified in *S. agalactiae* <SEQ ID 1015> which encodes the amino acid sequence <SEQ ID 1016>. Analysis of this protein sequence reveals the following:

```

Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1184 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 315

A DNA sequence (GBSx0344) was identified in *S.agalactiae* <SEQ ID 1017> which encodes the amino acid sequence <SEQ ID 1018>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 23
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4736 (Affirmative) < succ>
10     bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 316

A DNA sequence (GBSx0345) was identified in *S.agalactiae* <SEQ ID 1019> which encodes the amino acid sequence <SEQ ID 1020>. This protein is predicted to be elongation factor Tu (tufA). Analysis of this protein sequence reveals the following:

```

20     Possible site: 43
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3012 (Affirmative) < succ>
25     bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9737> which encodes amino acid sequence <SEQ ID 9738> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

30     >GP:BAB03851 GB:AP001507 translation elongation factor Tu (EF-Tu)
      [Bacillus halodurans]
      Identities = 302/397 (76%), Positives = 350/397 (88%), Gaps = 2/397 (0%)

35     Query: 7  MAKEKYDRSKPHVNIQIGHVDHGKTTLTAAITTVLARRLPSTVNPQKDYASIDAAPEER 66
      MAKEK+DRSK H NIQIGHVDHGKTTLTAAITTVLA+R V Y +ID APBER
      Sbjct: 1  MAKEKKDRSKTHANIGTIGHVDHGKTTLTAAITTVLAKNSGRGVAMA--YDAIDGAPBER 58

40     Query: 67  ERGITINPAHVEYETEKRYAHIDAPGHADYVKNMITGAACMDGAILVASTDGMPPQTR 126
      ERGITI+PAHVEYET+ RHYAH+D PGHADYVKNMITGAACMDG ILVV+ DGMPPQTR
      Sbjct: 59  ERGITITAHVEYETINRHYAHVDCPGHADYVKNMITGAACMDGGILVVSADGMPQTR 118

45     Query: 127 EHILLSRQGVKHLIVPMKVDLVDDBSLLELVMEIRDLLSEYDFPGDDLPVIQSGALK 186
      EHILLSRQGV +L+VF+NK D+VDDBSLLELVME+RDLLESDYDFPGDD+PVI+GSALK
      Sbjct: 119 EHILLSRQGVVPLVVF+LAKCMVDDBSLLELVMEVRDLLSEYDFPGDDVPVIRSGALK 178

50     Query: 187 ALEGDEKYEDIINELMSTVDEYIIPERDITKPLLLPVEDVFSITGRGTVASGRIDRTV 246
      ALEGD ++E+ I+ELM+ VD+YIP PERDT+KP ++PVEDVFSITGRGTVA+GR+RG +
      Sbjct: 179 ALEGDAEWEEKIELMAAVDDYIPTPERDTEKPPMPFVEDVFSITGRGTVATGRVERGQL 238

      Query: 247 RVNDEVEIVSIRKIQKAAVTVGVPMFRKQLDQSLAGDNVGVLLRGVQRDIERGQVLAQP 306
      V DEVEI+G++E+ +K VGVPMFRK LD AGDN+G LLRGV R+E+RGQVLAQP
      Sbjct: 239 NVGDEVEIIGLEEAQKTTVTVGVPMFRKLLVYEAAGDNIGALLRGVSRREVQRGVQVLAQP 298

```

-401-

Query: 307 GSINPHTRFKGEVYILSKKEGGRTTFFPNYRQPFYFRITTDVTSIELPAGTEMWPGDN 366
 G+I PHT FK RVY+LSKKEGGRTTFFPNYRQPFYFRITTDVTS I+LP G EAWWPGDN
 Sbjct: 299 GTITFHTNFKAIEVYILSKKEGGRTTFFPNYRQPFYFRITTDVTSIIQLPDGVMWVWPGDN 358

5 Query: 367 VTIEVELIHPIAVEQGTTFISREGGRTVSGSIVSEIRA 403
 V + VELI PIA+EGT FSIREGGRTVSG+V+ I+
 Sbjct: 359 VEMTVELIPIAIEBGTKFSIREGGRTVSGVWASIQ 395

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1021> which encodes the amino acid sequence <SEQ ID 1022>. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1367(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

20 Identities = 386/404 (95%), Positives = 396/404 (97%)

Query: 1 MEAPFKWAKEYDRSKPHVNIQTHGVHDGKTTTLTAITTVLARRLP+SVNQPKDYASID 60
 +EAPFKWAKEYDRSKPHVNIQTHGVHDGKTTTLTAITTVLARRLP+SVNQPKDYASID
 Sbjct: 12 LEAPFKWAKEYDRSKPHVNIQTHGVHDGKTTTLTAITTVLARRLPSSVQPKDYASID 71

25 Query: 61 AAPEERERGITINAHVSEYETEKHYAHIDAPGHADYVKNMITGAQMDGAILVWASTDG 120
 AAPEERERGITINAHVSEYET RHYAHIDAPGHADYVKNMITGAQMDGAILVWASTDG
 Sbjct: 72 AAPEERERGITINAHVSEYETATRYAHIDAPGHADYVKNMITGAQMDGAILVWASTDG 131

30 Query: 121 PMQTRHEIILSRQGVGKHLIVFMKVDLVNDEELLEVEIMEIRDLSEYDFPGDILFVI 180
 PMQTRHEIILSRQGVGKHLIVFMKVDLVNDEELLEVEIMEIRDLSEYDFPGDILFVI
 Sbjct: 132 PMQTRHEIILSRQGVGKHLIVFMKVDLVNDEELLEVEIMEIRDLSEYDFPGDILFVI 191

35 Query: 181 QGSALKALEGDEKVEDIIMELNSTVDEVIPEPERDTDKFILLPVEDVFSITGRGTVASGR 240
 QGSALKALEGD K+RDIIMRM TVD YIPEPERDTDKFILLPVEDVFSITGRGTVASGR
 Sbjct: 192 QGSALKALEGDTKFEIIMELNSTVDSYIPEPERDTDKFILLPVEDVFSITGRGTVASGR 251

40 Query: 241 IDRGTVRVNDEIVEIGIKEDIQKAVTVGVEMFRKQLDEGLAGDNVGVLLRGVORDEIERG 300
 IDRGTVRVNDE+IIVGIKE+ +KAVTVGVEMFRKQLDEGLAGDNVGV+LLRGVORDEIERG
 Sbjct: 252 IDRGTVRVNDEIIVGIKEIKETIKAVTVGVEMFRKQLDEGLAGDNVGVLLRGVORDEIERG 311

45 Query: 301 QVIAKPGSINPHTRFKGEVYILSKKEGGRTTFFPNYRQPFYFRITTDVTSIELPAGTEM 360
 QV+AKP SINPHTRFKGEVYILSK+EGGRTTFFPNYRQPFYFRITTDVTSIELPAGTEM
 Sbjct: 312 QVIAKPGSINPHTRFKGEVYILSKDEGGRTTFFPNYRQPFYFRITTDVTSIELPAGTEM 371

Query: 361 VMFGDNVTIIEVELIHPIAVEQGTTFISREGGRTVSGSIVSEIRA 404
 VMFGDNVTI VELIHPIAVEQGTTFISREGGRTVSGSIVSEIRA
 Sbjct: 372 VMFGDNVTINVELIHPIAVEQGTTFISREGGRTVSGSIVSEIRA 415

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 317

A DNA sequence (GBSx0346) was identified in *S.agalactiae* <SEQ ID 1023> which encodes the amino acid sequence <SEQ ID 1024>. Analysis of this protein sequence reveals the following:

55 Possible site: 36
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -0.64 Transmembrane 90 - 106 (90 - 106)

60 ----- Final Results -----
 bacterial membrane --- Certainty=0.1256(Affirmative) < succ>

-402-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 318

- 10 A DNA sequence (GBSx0347) was identified in *S.agalactiae* <SEQ ID 1025> which encodes the amino acid sequence <SEQ ID 1026>. This protein is predicted to be flsW. Analysis of this protein sequence reveals the following:

Possible site: 38
>>> Seems to have no N-terminal signal sequence
15 INTEGRAL Likelihood = -11.15 Transmembrane 44 - 60 (35 - 70)
INTEGRAL Likelihood = -4.73 Transmembrane 76 - 92 (74 - 98)
INTEGRAL Likelihood = -3.88 Transmembrane 117 - 133 (113 - 134)
----- Final Results -----
20 bacterial membrane --- Certainty=0.5458(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

25 >GP:AB39929 GB:U58049 putative cell division protein flsW
[Enterococcus hirae]
Identities = 78/159 (49%), Positives = 107/159 (67%), Gaps = 4/159 (2%)
Query: 1 MANSXYAMSGWNGPGRGLGNSIEKLGYLEATTDFVFSIVIEELGVIGAGFIALVFFLI 60
M+NS YA+ NGG FERG+GNSI K GYLPE+ TDF+FS++ RE G+IGA +L L+P L
30 Sbjct: 240 MNSYTYALYNGGLFGRGKNSITKGYLPESSTDFFSVIAEEPLGALLVLFLLFLLC 299
Query: 61 LRIMHVGIKXKDPFNSMIALGIMLLMQVFNIGCIGSLIPSTGVTFPPFLSQGNSLLV 120
+RI K K+ ++I +G+G +L+Q +NIG I GLIP TGV PF+S G3 S L+
35 Sbjct: 300 MRIFQKSTKQKQKQNLILIGVSTWILVQTSINTGSLIGLIPMTGVLPFVSTGTSYLI 359
Query: 121 LSAIGFVLINIDANEKELIMKRAEYKCFQENKELIN 159
LS AIG LNI ++ +E +++ + Q K K++N
Sbjct: 360 LSAIGLALNIBSRQKE---KNEQVERLQLKPKLLN 394

- 40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1027> which encodes the amino acid sequence <SEQ ID 1028>. Analysis of this protein sequence reveals the following:

Possible site: 51
>>> Seems to have no N-terminal signal sequence
45 INTEGRAL Likelihood = -10.93 Transmembrane 312 - 328 (303 - 338)
INTEGRAL Likelihood = -8.23 Transmembrane 22 - 38 (17 - 47)
INTEGRAL Likelihood = -6.85 Transmembrane 192 - 208 (187 - 211)
INTEGRAL Likelihood = -5.10 Transmembrane 218 - 234 (212 - 236)
INTEGRAL Likelihood = -4.83 Transmembrane 86 - 102 (85 - 107)
INTEGRAL Likelihood = -3.72 Transmembrane 385 - 401 (383 - 402)
50 INTEGRAL Likelihood = -3.45 Transmembrane 61 - 77 (61 - 79)
INTEGRAL Likelihood = -2.39 Transmembrane 344 - 360 (344 - 360)
----- Final Results -----
55 bacterial membrane --- Certainty=0.5373(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB59721 GB:AJ250603 FtsW protein [Enterococcus faecium]
 Identities = 131/397 (32%), Positives = 223/397 (55%), Gaps = 23/397 (5%)

5 Query: 15 KRHLNLYSILLYLILSVGLIMVYTSVGLIQANFPKSVINGVFWIISLVAITFI 74
 KR +++ IL PYL L6+IGL+ VYS +S L+QA N ++ Q +P +S I
 Sbjct: 3 KRKIDWILQPYLATSMIGLLEVVYSSVYRLQADENTKSLILRLQLIFLWSGVPLA 62

10 Query: 75 YKLKLNPLNTRVLTVMLEAFLLIAR--PPTTAIKNGHWIGVIGVSPQAPAYLKII 132
 +KL+L + + + F Li+ R P + CA NI + + FQE+
 Sbjct: 63 RSIKLHYLEPKIAGVGLAISIPFLILVRGIFGVTVNGAQWISLFGIQFQPSLEANLF 122

15 Query: 133 MVVYLALTFAKIQNISLYDQALTRKQWPTQNDLRDNRVYSILAVLVAQAQPLGNA 192
 +++YL+ F P +L+ + + + L+ QP + A
 Sbjct: 123 LIFVLSWFFRDGNN-----PPK--NLKPFLLTVSITLLILFQPKIAGA 164

20 Query: 193 SIIVLTAIMFISIGIGVWFSAILWMTGLSTVFLGTIAVIGVERVAKIP-VGVYAKR 251
 +I+ A ++F + + + +V + L G + G + +P +F + +R
 Sbjct: 165 LMLISIAMVIFMAAAVFPKGIYLVITFSALLIGAAGGVLYLGNK--GWLQPMFHAYER 222

25 Query: 252 FSAFFNFPHLDLDSGQILANSYYAMSGWFGQGLNSIEKRGVLEAQTDFVFSVIEE 311
 + +FF D +G+Q+ +S+YA+ NGS +G+GLNSI K+GYLE +TDF+FS+ EE
 Sbjct: 223 IATLRDPPIDSHGAGYQWTHSFYALNYOGIWRGLNSITKGYLFETETDFIFSITTE 282

30 Query: 312 LGLIGAGFILAVFFLLIRDMVIGIKAKNPFNMALGVGGMMALQVFNVTGIGISGLIPS 371
 LGLIGA +L L+P L +RI + + KN + IG G ++ +Q +N+G I+GL+P
 Sbjct: 283 LGLIGALCVLPLFLSLMRIFCLSSRCKNQAGLFLGFGTLLFVQTMNVGSIGLAFM 342

35 Query: 372 TGVTFPPLSQGNSLLVLSVAVGVFLNIDASEKRDDI 408
 TGV PF+S GG S L+L6+ +G INI + + + +
 Sbjct: 343 TGVPLPFVSYGOTSYLILSLGIGITINISSKIQREEL 379

An alignment of the GAS and GBS proteins is shown below:

Identities = 130/166 (78%), Positives = 152/166 (91%), Gaps = 2/166 (1%)

35 Query: 1 MANSXYAMNSGGWFGRLGNSIEKLGVLPEATTDFVPSIVIEKLGVIGAGFILAVFFLI 60
 +ANS YAMNSGGWFG+GLNSIEK GYLEA TDFVPS+VIEELG+IGAGFILAVFFLI
 Sbjct: 269 LANSYYAMNSGGWFGQGLNSIEKROYLPEAQTDFVPSVIEELGLIGAGFILAVFFLI 328

40 Query: 61 LRIMHVGIKAKDPFNSMIALGIGAMLMQVFNVIIGGISGLIPSTGVTFPFLSQGNSLLV 120
 LRIM+VGIAK+PFN+M+ALG+G M+LMQVFNVIIGGISGLIPSTGVTFPFLSQGNSLLV
 Sbjct: 329 LRIMHVGIAKKNPFNMALGVGGMMALQVFNVIIGGISGLIPSTGVTFPFLSQGNSLLV 388

45 Query: 121 LSVAVGVFLNIDASEKKELIKGAEBQYK--PQEKIEKILINDAPK 164
 LSVAV+GVFLNIDA+EK++ I KEAB Y+ +++N K+N+ F+
 Sbjct: 369 LSVAVGVFLNIDASEKRDIPKEALSYRKDTRKNSKVVNIKQPK 434

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

50 Example 319

A DNA sequence (GBSx0348) was identified in *S. galgaliae* <SEQ ID 1029> which encodes the amino acid sequence <SEQ ID 1030>. This protein is predicted to be probable cell division protein ftsW (ftsW). Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have an uncleavable N-term signal seq

55 INTEGRAL Likelihood = -9.77 Transmembrane 12 - 28 (7 - 37)
 INTEGRAL Likelihood = -7.22 Transmembrane 76 - 92 (74 - 97)
 INTEGRAL Likelihood = -6.53 Transmembrane 182 - 198 (178 - 201)
 INTEGRAL Likelihood = -4.62 Transmembrane 51 - 67 (46 - 69)
 60 INTEGRAL Likelihood = -2.87 Transmembrane 202 - 218 (202 - 218)

-404-

----- Final Results -----

5 bacterial membrane --- Certainty=0.4906(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9327> which encodes amino acid sequence <SEQ ID 9328> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

10 >GP:CAA44490 GB:X62621 ORF2 N-terminal [Lactococcus lactis]
 Identities = 82/199 (41%), Positives = 122/199 (61%), Gaps = 9/199 (4%)

Query: 1 MKIDKSHLLNYSILIPYLLISLGLVIVYSITSLIQLGAPFRSVINGQVFWAVSLVA 60
 M ++K + LNYSLIPYLLID+ +G+++I-STT +Q G NP++ VINQ F +S++
 15 Sbjct: 1 MNLKNNFNLYSILIPYLLISLGLVIVYSITSLIQLGAPFRSVINGQVFWAVSLVA 60

Query: 61 IIPYIKLNLNFKNSKVLTMVLEVFLLLIARF-----FTQEVNGAHQWIVIGPI-SF 113
 I IY+LKL LKN K++ + +++ + L+ R T VNGA GWI I I +
 20 Sbjct: 61 IAVIYRLKRLAKLNRMIGIIMVILLISLIPCRIMPSSFALPAPVNGARGWIIHPIGIGTV 120

Query: 114 QPAEYLKVIWYLAFTFARRQKKIETDYQALTGKWLFRSLDLKQWRFISLFMIGLV 173
 QPAE+ KV I+MILA F+ +Q+IE D + H+ L + L WR + + + +
 25 Sbjct: 121 QPAEFAKVFIIWYLAFTFARRQKKIETDYQALTGKWLFRSLDLKQWRFISLFMIGLV 173

Query: 174 IAQPDGLNYSIIVLTIVIM 192
 + PDGLN II +IM
 30 Sbjct: 179 LIMPDLNVTIIGAVALIM 197

There is also homology to SEQ ID 1028.

30 A related GBS gene <SEQ ID 8545> and protein <SEQ ID 8546> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 6
 MG0: Discrim Score: 15.18
 GVH: Signal Score (-7.5): -3.58
 35 Possible site: 34

>>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 5 value: -9.77 threshold: 0.0

INTEGRAL	Likelihood = -9.77	Transmembrane	12 - 28 (7 - 37)
INTEGRAL	Likelihood = -7.22	Transmembrane	76 - 92 (74 - 97)
40 INTEGRAL	Likelihood = -6.69	Transmembrane	210 - 226 (201 - 227)
INTEGRAL	Likelihood = -6.53	Transmembrane	182 - 198 (178 - 201)
INTEGRAL	Likelihood = -4.62	Transmembrane	51 - 67 (46 - 69)
PERIPHERAL	Likelihood = 1.32		116

modified ALOM score: 2.45
 45 *** Reasoning Step: 3

----- Final Results -----

50 bacterial membrane --- Certainty=0.4906(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

ORF02700[301 - 876 of 1377]
 55 BQAD[8615|8419(1 - 197 of 198) hypothetical protein in rpmg 3'region , fragment
 {Lactococcus lactis} SP|F27174|YAG2 LACIA HYPOTHETICAL PROTEIN IN RPMG 3'REGION (ORF2)
 (FRAGMENT). GP|44069|emb|CAA44490.1||X62621 ORF2 N-terminal {Lactococcus lactis}
 FIR|PC1134|PC1134 hypothetical protein 198 (xmpG 3' region) - Lactococcus lactis (fragment)
 %Watch = 15.1
 60 %Identity = 42.3 %Similarity = 64.9
 Matches = 82 Mismatches = 64 Conservative Sub.s = 44

-405-

[illegible]

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 320

25 A DNA sequence (GBSx0349) was identified in *S.agalactiae* <SEQ ID 1031> which encodes the amino acid sequence <SEQ ID 1032>. Analysis of this protein sequence reveals the following:

Possible site: 22
>>> Seems to have no N-terminal signal sequence

```

30      ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.3665(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

35 The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1033> which encodes the amino acid sequence <SEQ ID 1034>. Analysis of this protein sequence reveals the following:

Possible site: 54
>>> Seems to have no N-terminal signal sequence

```

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2373 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

Identities = 35/41 (85%), Positives = 37/41 (89%)

50 Query: 1 MEKEAKQIIDLKRNLFKIDVRAQKDEKVFMR TACCYSPPFY 41
+EKEAKQ+IDLKRNLFKIDVRAQKDEKVFMR TAC 8 Y
Subject: 1 LEKEAKQMIDLKRNLFKIDVRAQKDEKVFMR TACQSRVY 41

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-406-

Example 321

A DNA sequence (GBSx0351) was identified in *S. agalactiae* <SEQ ID 1037> which encodes the amino acid sequence <SEQ ID 1038>. Analysis of this protein sequence reveals the following:

```

Possible site: 49
>>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood = -1.65      Transmembrane      78 - 94 ( 78 - 95)
      INTEGRAL      Likelihood = -1.33      Transmembrane      421 - 437 ( 420 - 437)

----- Final Results -----
      bacterial membrane --- Certainty=0.1659(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15 >GP:CAA00827 GB:A09073 phosphoenol pyruvate carboxylase
      [Corynebacterium glutamicum]
      Identities = 335/958 (34%), Positives = 539/958 (55%), Gaps = 80/958 (8%)

20 Query: 22 EIITEVGLLKQLLDEATOKLIGSSPDKIE--KIVLSLITD---DYTLKETISALSNE 76
      + + + + L Q L E + G E + + E + + S + + L + +
      Sbjct: 3 DFLRDDIRFLQILGEVIAECBQGVETLVEQARLTSPDIARGNAEMDSLVCVQDGTIPA 62

      Query: 77 EMIVSRVYFSILPLINISEDVCLAYEINKNNINJOVLKSLT---TIDVV----- 125
      + +R FS LL N+ED+ Y L + L T T+D
      Sbjct: 63 KATPIARAFSHRALLANLAEL-----YBELRBOALDAGDTPDPSLDAITWLKINS 115

      Query: 126 -AGHENAKDILSRVNVVFLTAHPQVQRKIVLELTSKIHDLRKRYDRKAGIVNG--- 180
      G E D+L + V FVLTAHP+ +R+TV + I +R+ + +
      Sbjct: 116 NVGAERAVDVLRNAEAVFLTAHPTEIRRTKRVFDAQWITTHMRERHALQSAEFTARTQS 175

30 Query: 181 --EKVADLRVYIGIIMQTDITREKLVKNGEITVNGEITNRSILKAVTKLTAEYKALAA 238
      ++ ++R R I + QT IR + +++EI + Y Y SL++ + + +
      Sbjct: 176 KLDTEKNIRRRKITLWQTAIRVARPIEDIEVGLGKYKLLEIEIPRINRDAVAVEL 235

35 Query: 239 KK---GHLNFKPLTM-GHWIGGDRDGNFVTAETLRLSAMVQSEVINHYBIQMLLY 294
      ++ G+ L KP+ G WIGG DGNF+VTAET+ S +E++ +Y QL+L
      Sbjct: 236 ERFGEGVPL---KPVVKPGSWGIGDHDGNFVTAETVETGTHRAAETVLKYARGLHSL 292

40 Query: 295 RNSLSIMLTVSPSLVLTAMQSQGNSVYRNEPEYKAFNPTQDKLVCTLANLKVSSPK 354
      +SLG +V+P+L+ LA+ ++ R +EYR+A + + + + T
      Sbjct: 293 HELSLDRMKVTPQLLALADAGNDVPSRVDEPYRRAVGVAGRIAT----- 341

      Query: 355 EKFSVRQESSDITGVYKSHIAQVADIQTELPAYATAEEFKDILLVKGSLVYQGDS 414
      +++++G + + YN EEP D L + SL +
      Sbjct: 342 -----TAEIIGE-----DAVEGVW+KVTPYASPEEFLMDALTIDHSIRSKIVL 386

      Query: 415 LVDGEACLQIQVDINGFYLATIMRQDSSINEACVAELLSANIVDDYSSLSREKQQL 474
      + D L+ LI A+ PGF L +D+RQS E + EL + A + +Y LGE EK + +
      Sbjct: 387 IADRLSLVLSIAESPGFNILALDRQNSRYEDVLTPELRAQVYANRYELGRALKLV 446

50 Query: 475 LKELTEDPRTLSHAPKSELQKELIQTARELKDQLGHEDINQHLISHTESVDNF 534
      LKEL + + SE+ +EL IP+TA E + G + + IIS SV+D+
      Sbjct: 447 LKELASPRPLPHGSDRYEYTVRELGIPTASRAVKKPGPRMVPCHLSMSSVTDVL 506

55 Query: 535 ELAINLKEVGLIDN---QRIQIVLFTTIELDNLSDNEDIMQYLHYELVKWILATNN 590
      E + +LKE GLI AN + + +ELFTIEDL I+ + +L + + + +N
      Sbjct: 507 EPMVLLKEPGLIANGDNPRGTVDVILPFTYIEDLQAGAGILDELWKILDYRYLLQRDN 566

60 Query: 591 YQIMLGYSDSNKGGLYSSGWTLYKQQLKELIKIGRENGIKITTFPHRGGGTVVROGGGSPY 650
      QE+NLGYSDSNKGGLY S+ W LY A+ +L + + G+K+ PHRGGGTVVROGGGSPY
      Sbjct: 567 VQEVMLGYSDSNKGGLYPSANMALYDAELQVELCRAGVKKLRFPHRGGGTVVROGGGSPY 626

      Query: 652 EAITSPQFSIKDRIRLTQGEIIRNKYGNQDAAYYNLEMLSASTDRMVYNTINPHEI 710
      +AI +QP G++ +R+TQGEII KYGN + A NLE L+SA+++ + + +E+

```


-407-

Sbjct: 627 DAILAQPRGAVQGSVRITEQGEIISAKYGNFETARNLEALVSAILE---ASLLDVSEL 682
 Query: 711 DNFRFETMDGIVSESNV---YRNLVPMNPYPYDVPFASPIKEVSSINIGSRFAARKTT 766
 + + D I+SE + + Y +LV + + F DVF ++P++E+ SINIGSRP+RK
 5 Sbjct: 683 TDHQRAYD-IMSEISLSLAKYASLVHEDQGPFIDVYFTQSTPLQRIQINIGSRPSSRQQT 741
 Query: 767 TEISGLRALPWPVFSMSQNRMPGKYGVSAPKHEF---EQDEENLAKLQTMQYKMPFN 823
 + + LRAIPNV SMSQ+R+M PGN+GVG+A + I EQ +A+IGT+ + WPF
 10 Sbjct: 742 SSVEDRALPWPVLSMSQNRVMLPGVGVTALEQWIGBGBQATORIALQTINESWPFFT 801
 Query: 824 SLGSNDVMSLSMMIALQYQAGSKEVD-VEHILNEMQLTKMILATBQHMLR 882
 S+L N+ V+S K+ + A VA L EV + V++I E+ LTK M I D+L+
 Sbjct: 802 SVLDNMAQVMSKARLELAKLVADLIPUTEVARVRYSVIREZYFLTGGMFCVITGSDLLD 861
 15 Query: 883 ENFMGLASLDYRLPYFNVLMVQIELKLKLSNQLDEDYEKLHITTINGIATLRNSG 940
 +NP+L S+ R PY LW +Q+E++R R E + I +T+NG+T LRNSG
 Sbjct: 862 DNPLARSVQRYPYTLPLMIVQVBMRRYRKGDQSRQVSRNIQLTMGLSTALRNSG 919

A related GBS nucleic acid sequence <SEQ ID 10961> which encodes amino acid sequence <SEQ ID 10962> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1039> which encodes the amino acid sequence <SEQ ID 1040>. Analysis of this protein sequence reveals the following:

Possible site: 40
 >>> Seems to have no N-terminal signal sequence
 25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1613 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 30 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 659/927 (71%), Positives = 779/927 (83%), Gaps = 11/927 (1%)
 Query: 14 KLESSNKEIITEEVGLAKQLDEATQKLGSSFDKIRKIVLSLSITDDYTLGETISAL 73
 35 KLESS+N++II EEV LK++L+ T+++IG ++F IE I+ LS DY L++ ++ +
 Sbjct: 5 KLESSNQDIIEAEVALLKEMLENTIRMGDDAFTVIESIMVLSEQDYILEKVVANI 64
 Query: 74 SNSEMVIVSRYSFILPLINISESDVLAIRYNYKNINLQDYLGKLTSTIDVAGHENAKD 133
 SN+EM ++SRYSFILPLINISESDVLAIRYNY+N + DYLGKLT TI +AG +N KD
 40 Sbjct: 65 SNQEMEVISRYFSILPLINISESDVLAIRYNYQNNTDITDYLKGLTIDKLAGKMGKD 124
 Query: 134 ILRHVNVVPLTAHPTQVQRKTVLSLTSKHDLRLKRYRDVKGAGIVNKKQYADLRRIYGI 193
 ILE VNVVPLTAHPTQVQRKT+LELT+ IH LLKRYRD KAG++N EKK +L RYI +
 45 Sbjct: 125 ILRQVNVVPLTAHPTQVQRKTLSLTTIRHLKRLRYRDAKAGVINLKKQYELRYISM 184
 Query: 194 IMQDITREKKLVKNEITNVMEYTNRSLIKAVTKLTAEYKALAAKGIHLNPKPLTMG 253
 IMQDITREKKLVKNEITNVMYTYDGLIQAVTKLTETKYLKRLRYRDAKAGVINLKKQYELRYISM 244
 50 Sbjct: 185 IMQDITREKKLVKNEITNVMYTYDGLIQAVTKLTETKYLKRLRYRDAKAGVINLKKQYELRYISM 244
 Query: 254 NWIGDGRDGNPVTAEFTLRISAMVOSEVINHYISQINELYNRMSLSINLTVSPSLVTL 313
 NWIGDGRDGNPVTAEFTLRISAMVOSEVINHYISQINELYNRMSLSINLTVSPSLVTL 304
 55 Sbjct: 245 NWIGDGRDGNPVTAEFTLRISAMVOSEVINHYISQINELYNRMSLSINLTVSPSLVTL 304
 Query: 314 ANQSDNSVYRENEPYRKAFNFIQDKLVQTLNLAKVGSPPKKEFVSROESSDIVGRIKS 373
 A+ SQD S+YR NEPYR+AF++IQ +L QT + L + + SS + S
 60 Sbjct: 305 ASLSQDQSYTRGNEPYRAPHYIQRLKQTKQTQLT-----NOPASMSSSVGLNTAMS 358
 Query: 374 RIAQVADICTEELPAYATAREPKQDLLLVKQSLVQYQDLSVDGLACLIQAVDIPGFY 433
 A + + I AY + +FK DL +QSL+ G +L++G+L ++QAVDIPGF+
 65 Sbjct: 359 SPASLENPL-----AYDSVDPKADLKAIBGLDNGNLSALIGDLREVMQAVDIPGF 413
 Query: 434 LATIDMRQDSSINEACVARELLGANIVDQYSSLSEREKQQLLKELTDEPRTLSSTHAPK 493
 LA+IDMRQDSS+ EACVARELLK ANIVDQYSSLSEREK EKC +LL++L E+PRTLS K

Sbjct: 414 LASILMQDSSVQACVARKLLKGANIVDDYSSLSSTKKCDVLIQQIMERPRTLSSAAVAK 473

Query: 494 SEILLQKELATPQTARELKDQLGEDIINQHIISHTESVSDMPFLAIDMLKEVGLIDANQARI 553
S+LL+KELAI+ TARELKD+LGE++I CHIISHTESVSDMPFLAIDMLKEVGL+D +AR+

5 Sbjct: 474 SULLLEKELAIYTTARELKDQLGEEVIRQHIISHTESVSDMPFLAIDMLKEVGLDQQARARV 533

Query: 554 QIVPLFETIEDLNSRDIMTQYLHYLHVKKMIATNNYQEIIMLGYSDSNKGGLSSGWT 613
QIVPLFETIEDLDN+RDIM YL +++VK WIATN NYCEIMLGYSDSNKGGL+SGWT

10 Sbjct: 534 QIVPLFETIEDLDNARDIMAYLSHDIVKSIATNNYQEIIMLGYSDSNKGGLASGWT 593

Query: 614 LYKAQNELTKIGSEHGKIKITFFHGRGTVGRGGGSPYRAITSQPPGSIKDRIRITQZQEI 673
LYKAQNELT IGSE+G+KITFFHGRGTVGRGGGSPSY+AITSQPPGSIKDRIRITQZQEI

Sbjct: 594 LYKAQNELTAIGRHGVKIKITFFHGRGTVGRGGGSPYDAITSQPPGSIKDRIRITQZQEI 653

15 Query: 674 IENKYGNQDAAYNLEMLISASIDRMVITMITNPNIEDMFRSTMDGIVSSGNAYRNLVF 733
IENKYGN+D AYY+LEMLISASI+RMVT+MIT+PNEID+PKE MD IV++SN +YR LVF

Sbjct: 654 IENKYGNRDVAYYHLEMLISASINRMVITMITDPNIEDSFRIMDSIVADGNIYRKLVF 713

Query: 734 DNPYFYDVFPEASPIKEVSSLNIGSRPAARKTITEISGLRAIPWVFSMSQNRIMFPQWTG 793
DNP+FYDYFPEASPIKEVSSLNIGSRPAARKTITEI+GLRAIPWVFSMSQNRIMFPQWTG

20 Sbjct: 714 DNPYFYDVFPEASPIKEVSSLNIGSRPAARKTITEITGLRAIPWVFSMSQNRIMFPQWTG 773

Query: 794 VGSAPKHFTEQDEANLAKLQTYQKQPPFNLSLNVNVLKSNMNIQAQALQASKEV 853
VGSAPK+I++ +NL +LQ MYQ WPPF+SLSNVNVVLKSNMNIQAQAL+ +V

25 Sbjct: 774 VGSAPKRYIDRAQGNLERLQHWYQTFWFFSLSNVNVVLKSNMNIQAQALQAEQDV 833

Query: 854 RDVFNILIEBQCTKCNVILAIQDHDLLENPNMLHASLDYRLPYFNVLNVIQIELIKLR 913
RDVF IL+EBQCTK++ILAI+ HD+LLENP L SL RLPHYFNVLNVI+QIELIKR R

30 Sbjct: 834 RDVFYEILDEWQCTKCNVILAIQDHDLLENPNMLKSLKSLRFPYFNVLNVIQIELIKWR 893

Query: 914 SNQLEDYDKLIHTTINGIATGLRNSG 940
+NQLDE+ EKLIH TINGIATGLRNSG

Sbjct: 894 NNQLEDNDEKLIHTTINGIATGLRNSG 920

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 322

A DNA sequence (GBSx0352) was identified in *S. agalactiae* <SEQ ID 1041> which encodes the amino acid sequence <SEQ ID 1042>. This protein is predicted to be *Bacillus licheniformis* Pz-peptidase
40 homologue (pepF). Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3012 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1043> which encodes the amino acid
50 sequence <SEQ ID 1044>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3137 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

-409-

Identities = 512/593 (86%), Positives = 564/593 (94%)

Query: 1 MKLKKRSEFFPENELWDLTALYKDRQDFLLAIEKALBDIKVFKKNYEGKLNCEVDFTSALM 60
 5 Sbjct: 26 M+LKKRSEFFPENELWDLTALYKDRQDFLLAIEKAL+DI +FK+NYEG+L V+DFT AL+
 MELKKRSEFFPENELWDLTALYKDRQDFLLAIEKALQDIDLFKRNYEGRLTSVDFTQALDI 85

Query: 61 EIEHIYIQMSHIDTYAFMPQITDFSDNEEPAQISQAGSOFATKANVLLSFFWIALANADIK 120
 10 Sbjct: 86 EIEHIYIQMSHI TYAFMPQITDFSD+E PAQI+QAG DF TKA+V LSF+TALANAD+
 EIEHIYIQMSHIGTYAFMPQITDFSDSEPAQIQAQAGDFMTKASVALSFFDIALANADLD 145

Query: 121 ILDSLENNPHFKNTIRQAKIQKQHLSPVEKALINLNEVLNTPYDIYTRMAGDFDMGD 180
 15 Sbjct: 146 VLDTLEKNPFYSAAIRMAKIQKEHLLSPDVEKALANLREVINAOPYDIYTRMAGDFDMGD 205

Query: 181 FEVDGKTYKNSFVYIENFYQHENAIREK+FRSPFSGKLGKQNAAAAAYLAKVSEKLI 240
 20 Sbjct: 206 FEVDGKTYKNSFV+YEN++QHENAIREK+FRSPFSGKLGKQNA AAAAYLAKVSEKLI+
 FEVDGKTYKNSFVSYENFYQHENAIREK+FRSPFSGKLGKQNTAATAAYLAKVSEKLI 265

Query: 241 ADMRGYDSVDFYLLSBEQVDRSMDFRQIDILINDERQFVAQRFLGHADUNGIEKMTFADW 300
 25 Sbjct: 266 ADMRGYASVDFYLLABEQVDRSLDFRQIDILINTERGFPVAQRFLGHVAVUNGIEKMTFADW 325

Query: 301 KLDINDINLPEVSI+NDAYDLWKS+VAFGLKESYQVEYQKERMVDFAAANAKDSGGYAA 360
 30 Sbjct: 326 KLDINDINLPEVSI+NDAYDLWKS+VAFGLKESYQVEYQKERMVDFAAANAKDSGGYAA 385

Query: 361 DPKYKHPYVLMWSWGRMSDVYTLIHEIGHSQGFIPSDNHQSPFNTHMSTYTYVEAPSTFNE 420
 35 Sbjct: 386 DPKYKHPYVLMWSWGRMSDVYTLIHEIGHSQGFIPSDNHQSPFNTHMSTYTYVEAPSTFNE 445

Query: 421 LLLSDYLENQFDTPARQKRFALAHRLDTYFHNFTTHLEAA+PQRKVYTLIEGGTFGAQ 480
 40 Sbjct: 446 LLLSDYLEHQFDDPQKRFALAHRLDTYFHNFTTHLEAA+PQRKVYTLIEGGTFGAQ 505

Query: 481 LNAIMKEVLT+PKGDAI+IDDDAALYMMQAHNYMGLSYTYTYAGLVISTAGYINLKNP 540
 45 Sbjct: 506 LNAIMKEVLT+PKGDAI+IDDDAALYMMQAHNYMGLSYTYTYAGLVISTAGYINLKNP 565

Query: 541 NGAKEWLFLKSGGSRTPLETTALLISADISTDKPLRDTIPLNSIVDQII+Y+ 593
 50 Sbjct: 566 NGAKEWLFLKSGGSRTPLETTALLISADISTDKPLRDTIPLNSIVDQII+Y+ 618

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

45 Example 323

A DNA sequence (GBSx0353) was identified in *S. agalactiae* <SEQ ID 1045> which encodes the amino acid sequence <SEQ ID 1046>. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> May be a lipoprotein

50 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

55 The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1047> which encodes the amino acid sequence <SEQ ID 1048>. Analysis of this protein sequence reveals the following:

Possible site: 19

60

-410-

>>> May be a lipoprotein

----- Final Results -----

5 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below:

10 Identities = 72/127 (56%), Positives = 85/127 (66%)

Query: 1 MKCYIKLFLITVPATTLVACQFSTSNKTTSTLEWGVKVLVVKEDTNVLSKVVYHKG 60
 + K K L + A LVAC Q + +T T S V LVVKEDTN + KKV + KG
 15 Sbjet: 1 VNRKPKTGFLALVAMLLVACQGTGKIQITTSVFKADHHVRLVVKEDTNVDEKVPKGG 60

Query: 61 DTVLQVLKANYKVEKDGFTTSDIGISQDETKGLYWMFKVNNKLAAPKAANQIRVGGNDKI 120
 DTVLQVLK NY+VKEKDGFTT+IDGI QD YW+FKVN K+A K A+QI VK D I
 15 Sbjet: 61 DTVLEVLKKNYEVKEKDGFTT+IDGIBQDTKANKYVLFKVNKGADKQADQITVKGDSI 120

20 Query: 121 EFYQEVY 127
 EFYQEV+
 Sbjet: 121 EFYQEVF 127

25 SEQ ID 1046 (GBS185) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 28 (lane 6; MW 15.7kDa).

GBS185-His was purified as shown in Figure 199, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 324

30 A DNA sequence (GBSx0354) was identified in *S. agalactiae* <SEQ ID 1049> which encodes the amino acid sequence <SEQ ID 1050>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have a cleavable N-term signal seq.

35 INTEGRAL Likelihood = -4.46 Transmembrane 75 - 91 (67 - 94)
 INTEGRAL Likelihood = -4.41 Transmembrane 33 - 49 (30 - 49)
 INTEGRAL Likelihood = -2.60 Transmembrane 53 - 69 (52 - 70)
 INTEGRAL Likelihood = -1.38 Transmembrane 108 - 124 (106 - 124)
 INTEGRAL Likelihood = -0.06 Transmembrane 149 - 165 (149 - 165)

40 ----- Final Results -----

bacterial membrane --- Certainty=0.2784 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

45 A related GBS nucleic acid sequence <SEQ ID 9731> which encodes amino acid sequence <SEQ ID 9732> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10929> which encodes amino acid sequence <SEQ ID 10930> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1051> which encodes the amino acid sequence <SEQ ID 1052>. Analysis of this protein sequence reveals the following:

50 Possible site: 48
 >>> Seems to have a cleavable N-term signal seq.

-411-

```

INTEGRAL    Likelihood = -7.96    Transmembrane    50 - 66 ( 49 - 71)
INTEGRAL    Likelihood = -5.73    Transmembrane    101 - 117 ( 99 - 124)
INTEGRAL    Likelihood = -4.41    Transmembrane    141 - 157 ( 139 - 159)
INTEGRAL    Likelihood = -4.25    Transmembrane    73 - 89 ( 67 - 92)

```

----- Final Results -----

```

bacterial membrane --- Certainty=0.4185(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below:

Identities = 82/163 (50%), Positives = 120/163 (73%), Gaps = 3/163 (1%)

```

Query: 10  LTRVAILSALCVVLRVAFAPLPNIQIPITAIPLITVVLFDLKEGVATVITIMLVSSFLMGV 69
          ++R+AI+SALCVVLR F+ LFN+QF+TA L ++ F L E V ++ +S+FL+GF
Sbjct: 6  MSRIATMSALCVVLRVFPSSLPNVQPVTAFLSLYLLYFGLARAVLVMMLCLFLSALFLGF 65

Query: 70  GPVWFLQIISFTLILCLMKFLIYPLTKAVCPGKITEVVLQTFPAGGLGVVGVIIIDTCFA 129
          GPVWF Q+ F L+L LM+F++YPL++ F K ++ Q F G++YGV+IDTCFA
Sbjct: 66  GPVWFQVQVTCFVLVLLMRFFVLYPLSQQ--FPKY-QLGQQAFLNALGGLYGVLIIDTCFA 122

Query: 130 WLYHNPWWITYVLGSLFPMHAHALSTCLFYPLLILPILRRFRNEK 172
          +LY MPW++YVLG+ FN+AHALST +F+P+++ + RR E+
Sbjct: 123 YLYSNPWSYVLGMPFHIAHALSTLVFFVVMMLFRRLIGEQ 165

```

A related GBS gene <SEQ ID 8549> and protein <SEQ ID 8550> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1  Crend: 10
MoG: Discrim Score: 6.79
GVH: Signal Score (-7.5): -0.91
     Possible site: 28
>>> Seems to have a cleavable N-term signal seq.
ALON program count: 3 value: -4.46 threshold: 0.0
INTEGRAL    Likelihood = -4.46    Transmembrane    35 - 51 ( 29 - 54)
INTEGRAL    Likelihood = -1.38    Transmembrane    68 - 84 ( 66 - 84)
INTEGRAL    Likelihood = -0.06    Transmembrane    109 - 125 ( 109 - 125)
PERIPHERAL  Likelihood = 7.53      88
modified ALON score: 1.39

*** Reasoning Step: 3

----- Final Results -----
bacterial membrane --- Certainty=0.2784(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

ORF01220(421 - 552 of 1002)
GP|9950155|gb|JANG07353.1|AE004814.8|AR004814.16 - 56 of 69) hypothetical protein
{Pseudomonas aeruginosa}
%Match = 3.2
%Identity = 39.5 %Similarity = 60.5
Matches = 17 Mismatches = 15 Conservative Sub.s = 9

222      252      282      312      342      372      402      432
STLTKLTRVAILSALCVVLRVAFAPLPNIQIPITAIPLITVVLFDLKEGVATVITIMLVSSFLMGVGPVWFLQIISFTLIL
                                                                |:::
                                                                MDPSLFEEMHTGLTVILI
                                                                10

462      492      522      552      582      612      642      672
CLMKFLIYPLTKAVCPGKITEVVLQTFPAGGLGVVGVIIIDTCFAPLWLYHNPWWITYVLGSLFPMHAHALSTCLFYPLLILP

```

-412-

```

      : |::: | | :| || ||| | |
LFMAFTVVDLAKKSKAGKFGTLL--FFALGLGV-LGFTINGLVIGSLBAGAM
      30      40      50      60

```

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 325

A DNA sequence (GBSx0355) was identified in *S. agalactiae* <SEQ ID 1053> which encodes the amino acid sequence <SEQ ID 1054>. This protein is predicted to be endolysin. Analysis of this protein sequence

10 reveals the following:

```

Possible site: 28
>>> Seems to have a cleavable N-term signal seq.

```

- 15 ----- Final Results -----
- | | |
|-------------------------|--|
| bacterial outside --- | Certainty=0.3000 (Affirmative) < succ> |
| bacterial membrane --- | Certainty=0.0000 (Not Clear) < succ> |
| bacterial cytoplasm --- | Certainty=0.0000 (Not Clear) < succ> |

The protein has homology with the following sequences in the GENPEPT database:

- 20 >GP:CA72266 GB:Y11477 endolysin [Bacteriophage Baetille]
 Identities = 64/210 (30%), Positives = 95/210 (44%), Gaps = 15/210 (7%)
- Query: 66 KPIIDVSGWQLPKRIDYDTLSKNISGVVIRVFGSGKISKTNNAAYTTGIDKSPKTHKEF 125
 K I +D +E +ID+DT +S +R G + + +N +D+ +KT +
- 25 Sbjct: 12 KTIVDISHNA--DIDFDIAJNVSMFIARTDGHRYN--SNGELQGVDRKYATFVAMN 67
- Query: 126 QKRNIPVAVSYALGSSVSKMKRKAIFVYKNAAPYKPTFYWIDVSEETMSNNKGVQAFR 185
 + R I P Y + S V K+EA+ F+ N T + D E T NM + +Q F
- 30 Sbjct: 68 XARGIPPGNYMFRFSGVASAKQAEFFW--NYGDKDATVWVCDAEVSTAPNMKECICQFI 126
- Query: 186 KELKRLGAKNGVIGYTYFMTEQGISVIGVDFAVNIPTYGSDSGYYEAPQTELKYLHCY 245
 LK LGAK VG+YIG + E G D WIP YG+ + DL Q+
- 35 Sbjct: 127 DRLEKLGAKVGLYIGHHKYQKPGGKDVNCDFTWIPRYGNKFAF-----ACDLWQCN 177
- Query: 246 TSQGYLPGFNQPLDLNQLAVNKKKKTYRK 275
 T G + G + D+N + +K EK
- 35 Sbjct: 178 TEYGNIAIGIK-CDINVLGYDKPMSFTRK 206

- A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1055> which encodes the amino acid sequence <SEQ ID 1056>. Analysis of this protein sequence reveals the following:
- 40

```

Possible site: 31
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood =-16.98 Transmembrane 8 - 24 ( 3 - 28)

```

- 45 ----- Final Results -----
- | | |
|-------------------------|--|
| bacterial membrane --- | Certainty=0.7793 (Affirmative) < succ> |
| bacterial outside --- | Certainty=0.0000 (Not Clear) < succ> |
| bacterial cytoplasm --- | Certainty=0.0000 (Not Clear) < succ> |

- 50 An alignment of the GAS and GBS proteins is shown below:

Identities = 198/278 (71%), Positives = 235/278 (84%)

- Query: 1 MRRRIKPIVAVFSLFGLLLIIGHLSTNTLKKELVEAKKTIPSVKASKVPQKSTSSKD 60
 MRR+IKPIVV VFF L ++LIIG + + +KE+ +AK IP ++ K++S+
- 55 Sbjct: 1 MRRRIKPIVVLVFFILLAMVLIIGKRONHAKQKVEDAKSHIPLATSNPGKACTSTSET 60
- Query: 61 KEFVLKPIIDVSGWQLPKRIDYDTLSKNISGVVIRVFGSGKISKTNNAAYTTGIDKSPKT 120
 ++F+L PI+VSGWQLP+EIDYDTLS++ISG ++RV+GGS+I+ NNAA+TTGIDKSPKT

-413-

5
10
15

```

      Sbjct: 61  EDPIILNPFDVDSWGQLPEIDYDYLGRHIGCAIVRVYGGSGITAHNNNAAFTTGIDSKSPKT 120
Query: 121  HIKFQKRNIPFVAVYSYALGSSGVKMKEEAQIFPKMAAPKPTFFWIDVEETISMNMG 180
      HIKFQKRNIPFVAVYSYALG S KEMKESA FPKMAAP PT+XWIDVER TM +MKMG
      Sbjct: 121  HIKFQKRNIPFVAVYSYALGRSTKEMKSEARAFKQAAFYNTPTYYWIDVERATKDMKMG 180
Query: 181  VQAPRKLMLRLGAKNVGIYTYTPTBQSGISVGFDAWVITPYGSDSGTYRAAPQTELKY 240
      V AFR+ELK+LGA+NVG+YTYTPT BQ IS KGFDA+WVITPYGSDSGTYRAAP T L Y
      Sbjct: 181  VAFRBEELKGLGAKNVGLIYTYTPTMBADQISTGFGDSWIFPYGSDSGTYRAAPNTILDY 240
Query: 241  DLHQYTSQGYLGFPHQLDLNLQIAVNRDKKCTYKELPG 278
      DLHQYTSQGYL GFN LQLNLQIAV KD KKT+KKELPG
      Sbjct: 241  DLHQYTSQGYLSGFNNALDLNLQIAVTRDKTKTFFKELPG 278

```

A related GBS gene <SEQ ID 8551> and protein <SEQ ID 8552> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop Possible site: -1  Crend: 5
McG: Discrim Score: 13.20
GVH: Signal Score (-7.5): -0.72
Possible site: 28
>>> Seems to have a cleavable N-term signal seq.
ALOM program count: 0 value: 7.05 threshold: 0.0
PERIPHERAL Likelihood = 7.05 196
modified ALOM score: -1.91

*** Reasoning Step: 3

----- Final Results -----
bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

[illegible]

EAPACDLAQNTVEYGNLAGIGK-CDINVLVYGDKPMSPFTEKEGAKETLVPALNKVVTYEVGINLIPRIQDKLAFGLYEARI
 180 190 200 210 220 230 240

SEQ ID 8552 (GBS206) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 6; MW 31.7kDa).

GBS206-His was purified as shown in Figure 206, lane 6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 326

A DNA sequence (GBSx0356) was identified in *S. agalactiae* <SEQ ID 1057> which encodes the amino acid sequence <SEQ ID 1058>. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.44 Transmembrane 183 - 199 (183 - 200)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1574 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9729> which encodes amino acid sequence <SEQ ID 9730> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GF:AG320117 GB:AE005090 NADH dehydrogenase/oxidoreductase-like
 protein; Nola [Halobacterium sp. NRC-1]
 Identities = 38/156 (24%), Positives = 83/156 (52%), Gaps = 13/156 (8%)
 Query: 19 TMEILIAGSGSGLGKQIIKAALTGKHVAVLSRHGKGDIFKDPRLTVIRGSDITEADKIH 78
 +M++L+ GG+GF+G + + +GH V +R + D +T I GD+T + +
 Sbjct: 8 SMDVLVTGSGTGFTGHLCRELDGRGHEVTAFAREPADALEAD--VTRIVGDVTVKFEVA 65
 Query: 79 LEDRTFDLILDCIGA---IKPHOLD---ELNVKATQKAVALCHQIQIPKLVIYA--- 127
 D + + + + KP+ D + + + + T+ VA + + + + SA
 Sbjct: 66 NAIDGHDVAVNLVALSFLFKPSGGDSRHLDVHLGGTENVAVAASEAGVEYILQLSALLAD 125
 Query: 128 NSGYSAIYKSKRKABQIIKASGLDYLVFRPGLMYGE 163
 +G +AY+++K +AE+ ++S L + VRP +++G+
 Sbjct: 126 PTGPTAYLRAGKRAEAVRSSLHHTIVRPSVVFPG 161

No corresponding DNA sequence was identified in *S. pyogenes*.

A related GBS gene <SEQ ID 8553> and protein <SEQ ID 8554> were also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Crend: 5
 McG: Discrim Score: -7.99
 GvH: Signal Score (-7.5): -6.34
 Possible site: 41
 >>> Seems to have no N-terminal signal sequence
 ALOM program count: 1 value: -1.44 threshold: 0.0
 INTEGRAL Likelihood = -1.44 Transmembrane 183 - 199 (183 - 200)
 PERIPHERAL Likelihood = 4.29 20
 modified ALOM score: 0.79

*** Reasoning Step: 3

Example 327

A DNA sequence (GBSx0357) was identified in *S.agalactiae* <SEQ ID 1059> which encodes the amino acid sequence <SEQ ID 1060>. Analysis of this protein sequence reveals the following:

```

Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2850 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AA316953 GB:L23802 regulatory protein [Enterococcus faecalis]
Identities = 61/164 (37%), Positives = 96/164 (58%), Gaps = 13/164 (7%)

Query: 1  MSKKKKIKKTLDVQILDKAKIEH-----DSLQDLAQSDLPNGIQKQDIFKTLALI 51
      M+KK K+KKT++K+ +++++K+ + D L +++ L GI+K IFTKL +
Sbjct: 1  MAKKKTQKTNAMRMVEGHKVPYKEYEFAMSECHLSAESVAESL--GLEKGRIFKTLTV 58

Query: 52  QDKTGPPIIGILPLTEHLSEKKLAKISGNKKVQMIPQKDLQKITGYIHGANNPIGIRQKH 111
      G+KGTGP++ ++P + L KKLAK SGNKKV+M+ KDL+ TGYI G +P+G + K
Sbjct: 59  GNETGPFVAVIFGNQELDLKLAKLASGNKKVEMHLAKDLKLEATGYIRGGCSPTGM--KKQ 116

Query: 112 YPIFIIDTIALEKQELIVSAGEIGRSIRINSEVLADFNNAKFADI 155
      +P ++ A + +IVSAG+ G I + E + N +FA+I
Sbjct: 117 FPTYLAEEAQYSAIVSAGKRGWQILAPAILSLTNGQFAEI 160

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1061> which encodes the amino acid sequence <SEQ ID 1062>. Analysis of this protein sequence reveals the following:

```

Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2651 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 114/157 (72%), Positives = 139/157 (87%)

Query: 1  MSKKKKIKKTLDVQILDKAKIEHDSLQDLAQSDLPNGIQKQDIFKTLALIGDKTGPPIG 60
      M+KK K+KKT++K+QILDEA I H L+L+AL+GD P+ +Q DI+KTAL GD+TGP+IG
Sbjct: 1  MAKKKLKKKTLDVQILDKANTAHQGLKINALGSDPPDILQPSDIYKTLALTGDQTRPLIG 60

Query: 61  ILPLTEHLSEKKLAKISGNKKVQMIPQKDLQKITGYIHGANNPIGIRQKHNPYIFIDTIA 120
      I+PLTEHLSEK+LAK+SGNKKV M+PQKDLQK TGYIHGANNP+GIROKH+YPIPID A
Sbjct: 61  IILPLTEHLSEKQLAKVSGNKKVMVQKDLQKITGYIHGANNPVGIRQKHYSYPIPIDTA 120

Query: 121 LEKQELIVSAGEIGRSIRINSEVLADFNNAKFADIKK 157
      LEK ++IVSAGE+GRSI+I+S+ LADEFV A FAD+K+
Sbjct: 121 LEKQGIIVSAGEVGRSIRIKISSQALADFPVGSFADLKK 157

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 328

A DNA sequence (GBSx0358) was identified in *S.agalactiae* <SEQ ID 1063> which encodes the amino acid sequence <SEQ ID 1064>. Analysis of this protein sequence reveals the following:

-417-

Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 5 bacterial cytoplasm --- Certainty=0.4719 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8555> which encodes amino acid sequence <SEQ ID 8556> was also identified. This protein belongs to the glycolysis/gluconeogenesis pathway, and such proteins have been experimentally detected as surface-exposed in *Streptococci*. The protein has homology with the following sequences in the GENPEPT database:

- >GP:AAD36444 GB:AB001791 phosphoglycerate mutase [Thermotoga maritima]
Identities = 65/191 (34%), Positives = 93/191 (48%), Gaps = 13/191 (6%)
- 15 Query: 5 MKFYLVRHGKTCQNLBGRFQGGANGDSPLLEAIRELEELGQYLSSIHFDVAVYSSDLGRAR 64
 MK YL+RHG+T WN +G +QG D FL E E+ +L L + DA+YSS L R+
 Sbjct: 1 MKLYLIRHGGTITWNEKGLWQGVV-DVPLNERGRBQARKLANSKRV--DAIYSSPLKRL 57
- 20 Query: 65 DTVNILNDANS CPKEIHVTPQLRENALGTLHGCKIATMQAIYPRQMTAFYQNPLOPKHDM 124
 +T + A KEI LEE + G + YP + + +P M
 Sbjct: 58 ETAEEI--ARRFEKELIVEEDLRCEISLWNLGTLVREAIREYTFVEFKMGSSDP--NFGM 112
- 25 Query: 125 FGAESLYQTHRVESPLRSLSK---NYDKVLIVGHGANLTASIRSLGQYQSLHYKD 180
 G ES+ +RV + + S+ + V+IV H +L A I +LG LH
 Sbjct: 113 EGLESMPRVQNRVVKADHKIVSQEKLNGSENVIVSHSLSLRAFIQWILGLPL-YLRNF 171
- 30 Query: 181 KLDNASLTIIE 191
 KLDNASL+++E
 Sbjct: 172 KLDNASLVVE 182

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1065> which encodes the amino acid sequence <SEQ ID 1066>. Analysis of this protein sequence reveals the following:

Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 35 bacterial cytoplasm --- Certainty=0.3628 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
40 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 127/205 (62%), Positives = 152/205 (73%)

- 45 Query: 5 MKFYLVRHGKTCQNLBGRFQGGANGDSPLLEAIRELEELGQYLSSIHFDVAVYSSDLGRAR 64
 MK Y VRHGKT WN+GGRFQGA GDSPLLEA +E+ LG+ L+ + PDVAV+SDL RA
 Sbjct: 1 MKLYFVRHGKTLWNLBGRFQGGANGDSPLLEAKDEIHLGKELAKVAFDAVYTSILQRAM 60
- 50 Query: 65 DTVNILNDANS CPKEIHVTPQLRENALGTLHGCKIATMQAIYPRQMTAFYQNPLOPKHDM 124
 T I+ DA ++++T QLEH LG LEG KIATM AIYP+QM AF +N QPK D
 Sbjct: 61 ATAAIIIDPDQQPKLYHTDQLRENRGLKGAKIATMQAIYPPQMLAFRENLAQ+KPDQ 120
- 55 Query: 125 FGAESLYQTHRVESPLRSLSKNYDKVLIVGHGANLTASIRSLGQYQSLHYKDKLON 184
 F AES+YQCT RV ++S K+Y VLIVGHGANLTA+IRSLG+ L K LON
 Sbjct: 121 FEAESLYQTHRVCHLIQSFKDKIYQNVILVGHGANLTATIRSLGFEALLAKGLON 180
- 60 Query: 185 ASLTIETIEHDFKPKNLVNDKSYL 209
 ASLTI+ET D+ ++CL WNDKS+L
 Sbjct: 181 ASLTILETKDYLTIDCLVNDKSF 205

SEQ ID 8556 (GBS314) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 4; MW 27.2kDa), in Figure 169 (lane 15-17; MW 41.6kDa) and in Figure 239 (lane 4; MW 41.6kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 55 (lane 4; MW 52.1kDa).

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 329

A DNA sequence (GBSx0359) was identified in *S. agalactiae* <SEQ ID 1067> which encodes the amino acid sequence <SEQ ID 1068>. Analysis of this protein sequence reveals the following:

10 Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3014 (Affirmative) < succ>
15 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

20 >GF:CB12562 GB:Z99108 similar to hypothetical proteins [Bacillus subtilis]
Identities = 69/232 (29%), Positives = 108/232 (45%), Gaps = 9/232 (3%)

Query: 4 SIVFDVDDTIYDQAPYRIAEKCFDFMMSAINQAYIRFRHYSIDGPRVWAGWITSEY 63
+++FDVDDTI D QA +A+ F D ++ N +++ + + + G+ T +
25 Sbjct: 6 TLLFDVDDTILFDQAARALALRLLEDQKIFLINDMKAQKNTINQGLMRAFESGUTRDE 65

Query: 64 FRFRWCKETLLEFGYRIDEATGIYQRIYHELENITMLDEMNTLDFLKSQNVPM 123
R L E+GY EA G ++ Y LB L + L + + I+
25 Sbjct: 66 VVNTFRSALLKEYGY----EADGALLBQKYRFLFESGHLIDGAFDLISNLQQQDFLYIV 121

Query: 124 TNGPTHEQLKIVKKLGLYDVDPKRVISQATGPOKPEKTIENLAAEQF-DNNEFTLLYV 182
TNG + Q K++ GL+ + K + VS+ TGPOKP KE FN E+ + TL +
30 Sbjct: 122 TNGVSHTCYKRLRDSGLFPFF--KDI PVSEDGPOKPMKEYFNYPVPERIQPSAHTLLI 179

Query: 183 GDSVNDNDIMQAFQWGWMSWPNHGRSLKPGIKIPYVDVAIDNFEQLFGAVKV 234
GDS DI G G + WN + P I P Y+ I E+L+ ++
35 Sbjct: 180 GDSLTADYKGGLAGLDTCWNPDMKPNVPEIIPTYE--IRKLEELYHILNI 229

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1069> which encodes the amino acid sequence <SEQ ID 1070>. Analysis of this protein sequence reveals the following:

40 possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3216 (Affirmative) < succ>
45 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 276/300 (92%), Positives = 292/300 (97%)

50 Query: 1 MITSIVFDVDDTIYDQAPYRIAEKCFDFMMSAINQAYIRFRHYSIDGPRVWAGWITSEY 60
MIT+IVFDVDDTIYDQAPYRIA-EKCFDFMMS +NQAYIRFRHYSID+GPRVWAGWITSEY
Sbjct: 1 MITAIVFDVDDTIYDQAPYRIAEKCFDFMMSVWQAYIRFRHYSIDGPRVWAGWITSEY 60

55 Query: 61 TEYFRFRWCKETLLEFGYRIDEATGIYQRIYHELENITMLDEMNTLDFLKSQNVPM 120
TEYFRFRWCKETLLEFGYRIDEA G++PQE+YHELENITMLDEMNTLDFLKSQNVPM

Sbjct: 61 TTYFRPWRCKETLLEPGYREIDEAGVHPQEVYSHLENITMLDENRMTLDFLKSQNVPM 120

Query: 121 GIITNGPTEHQLKKVKLGLYDYDVPKRVISQATGQKPEKEIFMLAABQFDMNPSTTL 180
GIITNGPTEHQLKKV+KLGLYDY+D KRVISQATGQKPEKEIFMLAABQFDMNP TTL

5 Sbjct: 121 GIITNGPTEHQLKKVKLGLYDYDAKRVISQATGQKPEKEIFMLAABQFDMNPQTTL 180

Query: 181 YVGSYNDINDMGAPNGQWHSMMWPNHRGRSLKPGIKPVYDAIDNFBQLGAVKVLFDLPD 240
YVGSYNDINDMGAPNGQWHSMMWPNHRGR LKPG KPVYDAIDNFBQLGAVKVLFDLPD

10 Sbjct: 181 YVGSYNDINDMGAPNGQWHSMMWPNHRGRSLKPGIKPVYDAIDNFBQLGAVKVLFDLPD 240

Query: 241 NKPTFDINDKSNPVLGSLNGLMMAERLLSNMSVDKVVILLRLTAKQKVLRMKYAR 300
NKPTFD+NDK NP+L+MG+NGGLMMAERLLSNMS+DKVVILLRLT +QRKVL+KYAR

Sbjct: 241 NKPTFDVNDKKNPILQMGINNGLMMAERLLSNMSIDKVVILLRLTQKQKVLRLKYAR 300

- 15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 330

A DNA sequence (GBSx0360) was identified in *S.agalactiae* <SEQ ID 1071> which encodes the amino acid sequence <SEQ ID 1072>. Analysis of this protein sequence reveals the following:

- 20 Possible site: 34
>>> Seems to have no N-terminal signal sequence
- Final Results -----
- 25 bacterial cytoplasm --- Certainty=0.2451(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9727> which encodes amino acid sequence <SEQ ID 9728> was also identified.

- 30 The protein has homology with the following sequences in the GENPEPT database:

>GP: CAB11858 GB: Z59104 lysyl-tRNA synthetase [Bacillus subtilis]
Identities = 318/490 (64%), Positives = 390/490 (78%), Gaps = 1/490 (0%)

35 Query: 44 EELNDQIVRRERQALTEQGHIDPPGKRFERTATSGLANEKYADKSKDLHDIESTATIA 103
EELNDQ VRR+RM L + GIDPRG RFERT S ++ Y D +HE+L + TIA
Sbjct: 9 EELNDQLQVRDKQWGLRDNQIDPPGARFERTHQQEVISAYQDLTKEELKEKAIEVTTA 68

Query: 104 GLRATKRGKGGKGFAGHIDQREQIQIYVRKDSVGENYEIPFKADLQDFLGVGRQWRTD 163
GR+NTKRGKGGK GFAG+QD EQIQIYVRKDSVG+ YEIPK +DLQD +GV G+V +T+
40 Sbjct: 69 GQWNTKRGKGGKGFAGHLDLQEQIQIYVRKDSVGDQYEIPKSDGLDGLGVGKVPKFN 128

Query: 164 MGELGKATHITHLGKARLPLEKPHGLTDIETTYRPHLDLIGNRDSFORFVTRSKIIIS 223
+GELG+KAT L+KALRPLP+K+HGL D+E YR+R+LDL N DS F+TRSKII
45 Sbjct: 129 VGELSVKATSFELLTKALRPLPDXYHGLKDVEQRYRQRYLDLWNPDSEKTFITRSKIIQ 188

Query: 224 EIRRFMDGNGFLVSTPVLNENAGGASARPPITTHNMQIDMVLRITATELHLKELIVGM 283
+ER+D +G+LEVETP +H+ GGASARPPITTHNA DI + +RIA ELHLKELIVQD+
50 Sbjct: 189 AMRYYLDDGKYLEVETPTMHISIPGASARPPITTHNALDIPVMRIAELHLKELIVGSL 248

Query: 284 ERVYEIGRIIPRNGEMDATHNPEFTSIRAYQAYADYQIDMOLTSGLIQHVTKTVGDEGPDN 343
E+VYEIGR+PRNEG+ HNPEFT IE Y+AYADY+DIM LYE ++ H+ + V G I
55 Sbjct: 249 EKVYEIGRVPRNEGVSIRHNPEFTMIELYEAYADYKDIMSITENLVAHIAEQVILGTTTI 308

Query: 344 YCOTEIKINEPFRKVRHMDAVKTEGTIDPKKMTLEQAQALACQKNVPLEKHPTTVGHII 403
Y + I + +KR+HMDVAVKE TG+DPW+E+T+B+A+ A+E V + K TVGHII
60 Sbjct: 309 YGEEQIDLKPEKRIIMVDAVKEATGVDPKKEVTVGQAREYAKHEHEVEI-KDSMTVGHII 367

Query: 404 NAFPEFVEDTLIQPTFFVFGHVPKVSPLAKKNIDTPRPTDRPELFINTKEYANAFTELND 463
N FFE+ +E+TLIQPTF++GHPVBS+PLAKKN DTPRPTDRPELFINTKEYANAFTELND
Sbjct: 368 NEFPQKRIETLIQPTFFYGHVPVEISPLAKKNIDTPRPTDRPELFIVGREHANAFTELND 427

-420-

Query: 464 FIDQLSRFEAQASAKELGDDEATGVDDYVEALEYGMPPPTGGLGIGIDRLCMLLTDTTTI 523
 PIDQ RFEAQ +E G+DRA +D D+VEALEYGMPPPTGGLGIGIDRL MLIT+ +I
 Sbjct: 428 PIDQRERFEAQLKEREAGNDRAHLMDEDFVEALEYGMPPPTGGLGIGIDRLVMLLTNAPSI 487

Query: 524 RDVLLFPPTMK 533
 RDVLLFP M+
 Sbjct: 488 RDVLLFPQMR 497

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1073> which encodes the amino acid sequence <SEQ ID 1074>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4694(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 439/500 (87%), Positives = 474/500 (94%)

Query: 34 LEEIMSNQHIEELNDQIVRREKMAALTEQGIDPPGKRFPRTATSGQINBKADKSKEDL 93
 LEE MSNQHIEELNDQIVRREKMAAL EQGIDPPGKRFP+KTA S +D EKYADK+KE+L
 Sbjct: 1 LEEIMSNQHIEELNDQIVRREKMTALASQIDPPGKRFPDRTANSABLKADKSKEDL 60

Query: 94 HDISETATIAGRMTKRGKGVFAHQDREGQIQIYVRKDSVGBENYEIPKADLGDFF 153
 H++ ETA +AGRLMTKRGKGVFAH+QDREGQIQ+YVRKDSVGB+NYEIPKADLGDFF+
 Sbjct: 61 HELNETAVAGRMTKRGKGVFAHLQDREGQIQIYVRKDSVGBENYEIPKADLGDFF 120

Query: 154 GVBGQVMRTDWSLSIKATHITHLSKLRPLPEKFGHGLTDIETIYRKRHLDLISNRDSD 213
 GVBG+VMRTDWSLSIKAT +THLSK+LRPLPEKFGHGLTDIETIYRKRHLDLISNR+SD
 Sbjct: 121 GVBGQVMRTDWSLSIKATKLTLSKLRPLPEKFGHGLTDIETIYRKRHLDLISNRSD 180

Query: 214 RFVTRSKIISEIRPMDSNGFLEVETPVLNHEAGGASARPPTTHNQAQIDMWLRATL 273
 RFVTRSK+ISEIR+D+ FLEVETPVLNHEAGGASARPPTTHNQAQ+IDMWLRATL
 Sbjct: 181 RFVTRSKIISEIRRYDGLDFLEVETPVLNHEAGGASARPPTTHNQAQIDMWLRATL 240

Query: 274 HLKRLIVGGMERVYEIGRIFRNEGDATNDEPFTSIEAYQAYADYQIDMLTSGIHT 333
 HLKRLIVGGMERVYEIGRIFRNEGDATNDEPFTSIE YQAYADY DIW+LTSGIHT
 Sbjct: 241 HLKRLIVGGMERVYEIGRIFRNEGDATNDEPFTSIEVYQAYADYDIDMLTSGIHT 300

Query: 334 KTVGQGPINYGQTEIKINEPKRVHVDVAKEITGIDFWKEMTLEEAQALAEKNVPLE 393
 K V+GQGP+YQGTET+INEPKRVHVDVA+KE+TG DFW EMT+SEA ALA+EK VPLE
 Sbjct: 301 KAVRGQGPIDYQGTETINEPKRVHVDVAKEITGADFWPEMTVEEAALAEKQVPLE 360

Query: 394 KHFSTTVGHINAFPEEFVEETLQPTFVFGHPVEVSLAKKNDTRDPTDRFLPTINTKE 453
 KHF +VGHINAFPEEFVE+TL+QPTFVFGHPVEVSLAKKNDTRDPTDRFLPTINTKE
 Sbjct: 361 KHFISVGHINAFPEEFVEETLVQPTFVFGHPVEVSLAKKNDTRDPTDRFLPTINTKE 420

Query: 454 YANAFTELNDPIDQLSRFEAQASAKELGDDEATGVDDYVEALEYGMPPPTGGLGIGIDRL 513
 YANAFTELNDPIDQLSRFEAQAKELGDDEATGVDDYVEALEYGMPPPTGGLGIGIDRL
 Sbjct: 421 YANAFTELNDPIDQLSRFEAQAKELGDDEATGVDDYVEALEYGMPPPTGGLGIGIDRL 480

Query: 514 CMLLTDTTTIRDVLLFPIMK 533
 CMLLT+TTTIRDVLLFPIMK
 Sbjct: 481 CMLLTNTTTIRDVLLFPIMK 500

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-421-

Example 331

A DNA sequence (GBSx0361) was identified in *S.agalactiae* <SEQ ID 1075> which encodes the amino acid sequence <SEQ ID 1076>. This protein is predicted to be 6,7-dimethyl-8-ribityllumazine synthase (ribH). Analysis of this protein sequence reveals the following:

5 Possible site: 34
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.1042 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

15 >GP:CAE14257 GB:Z99116 riboflavin synthase (beta subunit) (Bacillus subtilis)
 Identities = 103/151 (68%), Positives = 120/151 (79%)

 Query: 1 MTIIEGQLWANEMKIGIVVSRFNEIITSKLLSGAVDGLRHGVSEEDIDIVVWVGAPEIP 60
 M II+G LV +KIGIVV RFN+ ITSKLLSGA D LIRHGV DID+ WVGAFPEIP 60

 Sbjct: 1 MTIIQGEHLVGTGLKIGIVVGRFNDFITSKLLSGAEDALLRHGVDTNDIDIVWVGAPEIP 60

20 Query: 61 YMARFGALYKDYDAIICLGVIKSGTDHYDYVCNEVTKIGHLSNQSODIPHIFGVLTIDN 120
 + A+KGA K YDAII LG VI+G+T HYDYVCNE KGI + + P IFG++TT+N

 Sbjct: 61 FAKAKGATYKKYDAIITLGTIVGATTHYDYVCNEAAKGIAQAANTTGVVPVIFGVITEN 120

25 Query: 121 IEQAIERAGTKAGNKGYDCALSAIEMVNLDK 151
 IEQAIERAGTKAGNKG DCA+SAIEM NL++

 Sbjct: 121 IEQAIERAGTKAGNKGVDCAVSAIEMANLAR 151

No corresponding DNA sequence was identified in *S.pyogenes*.

30 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 332

35 A DNA sequence (GBSx0362) was identified in *S.agalactiae* <SEQ ID 1077> which encodes the amino acid sequence <SEQ ID 1078>. This protein is predicted to be GTP cyclohydrolase II (ribA/B). Analysis of this protein sequence reveals the following:

 Possible site: 20
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.1918 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

45 A related GBS nucleic acid sequence <SEQ ID 9725> which encodes amino acid sequence <SEQ ID 9726> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

50 >GP:AAA86524 GB:U27202 GTP cyclohydrolase II/
 3,4-dihydroxy-2-butanone-4-phosphate synthase
 [Actinobacillus pleuropneumoniae]

 Identities = 230/395 (58%), Positives = 307/395 (77%)

 Query: 19 FSPKIKLLQDIKSGKVVLMDDENRENKGDLCIAEMVTKESINFMKFGKGLICLPLSN 78
 FS ++ ++ I+ GK++++ DDE+RENKGD ICAAE T E+INPMA +GKGLIC P+S

 Sbjct: 6 FSKVEDAEIAIRQGKIIIVTDDEBRENKGDFTCAARFATPENTINFMATYKGLICTPIST 65

Query: 79 YYAEKLELAQMSHNTDNHETAFITSIDHLSTSTGISAEDRALTAQVANDSSKADFER 138
 A+KL M + N DNHETAFIT+SDH+ T TGISA +R+TA + +D+KA DFER
 Sbjet: 66 ETAKKLNPHMVAQNDHETAFITVSDHIDTGTGISAFAERSITAMKIVDNDKATDFER 125

Query: 139 PGHLPFLAKEGGVILRNHGTETATVULCRILAGLKEGLOCEIMAEQSGMNRKIDELLAFAP 198
 PGI+PFL+AKEGGVIL RNHGTETATVUL RLAGLK GLCEIMA+DG+NM +L FA
 Sbjet: 126 PGHLPFLAKEGGVILRNHGTETATVULARILAGLKHAGLCEIMADGDMNMTPDLQKFAV 185

Query: 199 KHDLAIAITIKQLQDYRQKSGGGVREIKIQLPTQPGHMTAYGYSEVVANKRHVALVKGDI 258
 +H+ TI+QLQY+YR+ + V + +++PT+G F A+ + EV+ KBIHVALVKGDI
 Sbjet: 186 EHNMPFITTIQQLQYRKRKDSLVKQISVVVMPTKYGRPMHNS FVIVLSGKRHVALVKGDI 245

Query: 259 SSGEDVLCKLHSECLTGDVFSHSLRCDGSGQLANALQTEAEGRGVLYMNRQSGRGIGLIN 318
 + GE VL R+HSECLTGD F S RCDG+Q A A+ QTE EGRGV+LY+RQGRGIGLIN
 Sbjet: 246 TDGSQLARITHSECLTGDAPGSGQRCDGQQAAMTQTEQEGRGVILYLQGRGIGLIN 305

Query: 319 KLRAYHLQEGGLDTLEANLALGPEGDERDYGVSAQLLDKLGINSINLITNMPDKIQLEA 378
 KL+AY LQ++G+DT+SAN+ALGF+ DER+Y + AQ+ + LG+ SI LITNMP KI+ L+
 Sbjet: 306 KLRAYELQKGMUTVEANVALGFIORDEREYTYGAQMPQQLGVKSIRLLITNMPAKIEGLKE 365

Query: 379 BGICVYKRVPLQAVTAZYDLNLYLTKKRMCHLLD 413
 +G+ + R P+ V D++YLK K+ RMGH+ +
 Sbjet: 366 QGLNVAREPITIVPNNKIDILYLVKQIKRMGMFN 400

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 333

A DNA sequence (GBSx0363) was identified in *S.agalactiae* <SEQ ID 1079> which encodes the amino acid sequence <SEQ ID 1080>. This protein is predicted to be riboflavin synthase alpha chain (ribE). Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.3517 (Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000 (Not Clear) < succ>
bacterial outside ---	Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9723> which encodes amino acid sequence <SEQ ID 9724> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB05274 GB:AP001512 riboflavin synthase alpha subunit [Bacillus halodurans]
 Identities = 98/216 (45%), Positives = 147/216 (67%), Gaps = 2/216 (0%)

Query: 1 MFTGIIERMQQVSRIRNGIKSQSLIDAPKLVPLLRKGDVAVNGVCLTVLDRKSETAFIA 60
 MFTGIIIE+G + I+ ++ ++I + K+V ++ GDS+AVNGVCLTV ++T F
 Sbjet: 1 MFTGIIEDVGTIDAIQQYGEALVMTITSKKIVSDVQLGDSIANGVCLTVTSTPTDQFTV 60

Query: 61 DVMPESSMRTSLAALRLHKNVLEALRSDRLGHPVLGHVDGKVEIKQDDIAVRF 120
 D+KPE++ TSL L S+VNL E A+ ++ R GGH V GHVDG+G I K ++ D AV +
 Sbjet: 61 DLMPTVTRATSLRLLEKSGRVNLERAMVANGRFGGHIVSGHVDGIGITIRKKERKNQVTV 120

Query: 121 SIDAPPSIMSYIIEKSGVALDGISLTVVSTFHSPEVSVIPHTMAQTNLKLKVGDLINI 180
 +I+ S+ Y+I KGSVA+DG SLT+ ++ +F +S+IPHTM +T + LKX GD+NI
 Sbjet: 121 TIEVSSSLRRYMIHKGSAVDGTSITLTFDVSDDKFTPTISITPHMETTIGLKKAGDIVNI 180

-423-

Query: 181 EVDVLGKYAEKFLAPINRNTHTSSVMDWSFLSENGY 216
 E D++GKY E+F+ N + +F++E+GY
 Sbjct: 181 BCDLIGKYIEQFVQGGKPKVNEGG--LTKAPLITEHY 214

- 5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 334

- A DNA sequence (GBSx0364) was identified in *S.agalactiae* <SEQ ID 1081> which encodes the amino acid sequence <SEQ ID 1082>. This protein is predicted to be riboflavin-specific deaminase (ribD). Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.01 Transmembrane 307 - 323 (307 - 323)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1404 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA86522 GB:U27202 riboflavin-specific deaminase [Actinobacillus pleuropneumoniae]
 Identities = 182/353 (51%), Positives = 259/353 (72%)
 Query: 6 DYMALALKEAKMGFVAFNPLVGVIVIKDRIISKGYHFRFGDLHAERGAIRNADEDIS 65
 DYM A+ A++G+G+ ENPLVG VIVE+ I++GYH++ G HAER A+ + ED+S
 Sbjct: 51 DYMRAALALAKGLGWTNPNPLVGCIVIRNGEIVABGYHEKIGGWHAEARNVHLCKEDLS 110
 Query: 66 GSTLVITLPECCCHVGKQPPCTEALIKSGIKVIVVSGSDPNPLVSGKGIALLRKGBLAVIEV 125
 G+T YVTLPECC G+ PPC++ LI+ GLIKV +GS DPNPLV+G+G LR+ G+ V
 Sbjct: 111 GATAYVTLPECCCHGHTPPCSDDLTERGIGKVKVFIGSSDPNPLVHGEGANQLRQAGVEVVE 170
 Query: 126 GILREECDAINERFIHFHYTKQFPVYLKYMATLDGKIATKTGDSKNWISNEHSRQSVQKLR 185
 G+L+ECCDAIN F ++ K+P+V +KYMT DGIAT +G+SKWT+ E +R VQ+ R
 Sbjct: 171 GLKKECCDALNPIFFHYIQTKRPVYLKYMATADGKIATGSGESKNITGESARARVOOTR 230
 Query: 186 OKCSAINVGINTVLADNPRITCRIPKGBALVRIVCDGSQLKPLDSYLVKSAKTITPMIAT 245
 +SAINVG++TVLADNF L R+P + VRIVCDGSQL+ PLD LV++AK T IAT
 Sbjct: 231 HQYSAINVGVDTVADNFMNLSRMPNAKQFVRIVCDGSQLTPLDQCLVQTAKEYRTVIAT 290
 Query: 246 CSDMIAQQCTIAKMGCRLLIKVPRKDGKLDLKVMTILQBGIDSLITGGSGSLHPSALKA 305
 SD+L + + + +G ++ ++ ++DL+ L LG+ IDELL+EGGSSL+PSAL++
 Sbjct: 291 VSDDIQRIQRFPLGVVVLVCKARKNRVLDQLQLKLGEMQIDSLILEGSSSLNPSALES 350
 Query: 306 GIVNRILVETAPKIIIGGLAKTALISGEGDLMLANQAFKVIDELGRWDSQVIE 358
 GIVNR+ +IAPK++GG +2KT I GEG+ ++QA ++K + D++++
 Sbjct: 351 GIVNRVHCYIAPKLVGGRQAKTPIGGE3IQQIDQAVRLKLGKSTRLIGEDILLD 403

- 50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1083> which encodes the amino acid sequence <SEQ ID 1084>. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.17 Transmembrane 88 - 104 (88 - 105)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1468 (Affirmative) < succ>

-424-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

5 >GP: CAB11794 GB:299104 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 71/161 (44%), Positives = 109/161 (67%)

Query: 13 LEEQTFPMQREALKEAKSLQKAEIPICVIVKDGRIIGRGNHAREESNQAIMHAEINMAIN 72
 + + M+EA+KEA+K+ +K E+PIG V+V +GEII R HN RE ++I HAE+ I+
 10 Sbjct: 1 MTQDRLYMKEAIKKAKEAKGKVPITGAVLVINGEIIARAHNIRETEQRSIAHAEMLVID 60

Query: 73 EANAHEGNWRLDITTLFTVETPCVMCSGAIGLARIPIHVIYASNQKFGVDLSYLQILTDE 132
 EA G WRL TL+VT+EPC MC+GA+ L+R+ V+GA+ K G +L +L +E
 Sbjct: 61 EACFALGTWRLEGATLYVTLEPCFMCGAVVLSRVEKVVFGAFDPKGGCSOTLWNLQEE 120

15 Query: 133 ELNHRVQVERGLLAADCANIMQTFPRQGRERKIAKHLIKE 173
 R NH+ +V G+L +C ++ FFR+ R++KK A+ + E
 Sbjct: 121 RFNHQAEVSVGLSEECGMLSAFFRELKKKKKARCNLSE 161

20 An alignment of the GAS and GBS proteins is shown below:

Identities = 48/146 (32%), Positives = 71/146 (47%), Gaps = 21/146 (14%)

Query: 7 YMALAKAEAKGKMGFVAFNPLVGAVIVKDDRIISKGYHKRFGD---LHAERQAIKNADE 62
 +M ALKEAKK + A P +G VIVKD II +G+ R +HAE AI A+
 25 Sbjct: 19 FMQEALEAKESLQ-KAEIP-IGCVIVKDGRIIGRGNHAREESNQAIMHAEINMAIN 76

Query: 63 D-----ISGSTLYVTLEPCCHVGKQPCTEALIKSGIKKVVVGSIDPNPLVSGKIALLR 117
 + +TL+VT+EPC C+ A+ + I V+ G+ + +L
 Sbjct: 77 HEGNWRLLDITTLFTVETPCV-----MCSGAIGLARIPIHVIYASNQKFGVDLSYLQILT 130

30 Query: 118 KEGLN----VEVGILREECALNERF 139
 E LN VE G+L +C + + F
 Sbjct: 131 DERLNRHVQVERGLLAADCANIMQTF 156

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 335

A DNA sequence (GBSx0365) was identified in *S.galactiae* <SEQ ID 1085> which encodes the amino acid sequence <SEQ ID 1086>. This protein is predicted to be Nramp metal ion transporter. Analysis of this
 40 protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

45 INTEGRAL Likelihood = -11.89 Transmembrane 169 - 185 (160 - 191)
 INTEGRAL Likelihood = -11.09 Transmembrane 140 - 156 (128 - 165)
 INTEGRAL Likelihood = -6.85 Transmembrane 359 - 375 (354 - 379)
 INTEGRAL Likelihood = -6.48 Transmembrane 269 - 285 (263 - 287)
 INTEGRAL Likelihood = -6.16 Transmembrane 426 - 442 (423 - 445)
 INTEGRAL Likelihood = -5.57 Transmembrane 62 - 78 (58 - 80)
 INTEGRAL Likelihood = -4.94 Transmembrane 107 - 123 (103 - 127)
 50 INTEGRAL Likelihood = -4.46 Transmembrane 391 - 407 (389 - 408)
 INTEGRAL Likelihood = -4.35 Transmembrane 310 - 326 (307 - 328)

----- Final Results -----

55 bacterial membrane --- Certainty=0.5755 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF63825 GB:AE003939 manganese transport protein [Xylella

-425-

fastidious]
 Identities = 192/436 (44%), Positives = 274/436 (62%), Gaps = 14/436 (3%)

5 Query: 10 SLSEVNQSVVEPHNSSFNNTLRAFLGPGALVAVGYMDPGNWTISVIGGATYRLLLPVL 69
 SL R+ SV V + L AFLGEG +V+VGYMDPGN T + GG+ Y+LL V+L
 Sbjct: 39 SLGEMHSAVASVRGHWGPRLLAFLGPGTVMVSVGYMDPGNWTAGLAGSSRPGMYLLSVTL 98

Query: 70 VSLIMAMQIQQAGKIGIVTRQDLAQATASRLPKPKRYLLFIIRIALIATDGLAEVIGSA 129
 +S++MA+ IQ +A +LGI + DLQA +R + L+++ ELR+IA DLAEVIG+A
 10 Sbjct: 99 LSNVMAIVLQALARLGIASMDMLAQACRARSRTGTHALWVVCLELATIACDLAEVIGTA 158

Query: 130 IALHLLFGWPLLSIMITITLDVFLLLMLKLGKIRAFVSVLIITILIPYLVVLSQF 189
 IAL+LL G F++ ++IT +DV L+LLM G + +BAFV L+L I F +VI+ P
 15 Sbjct: 159 IALNLLGVPIIWGVVITADVVLVLLMERGPRALEAFVIALLLVPGCVQIVLAAP 218

Query: 190 DLDAMFGFLPHEHELPNISHBESKNSPLTLAIGIGATVMEHNLVHSSLSQTRRVDYHNK 249
 L + GF+P ++ L DA+GI+GATVMEHNLVHSS+ QTR
 Sbjct: 219 FLQEVLGSGFVRWQV-----ADPQALYLAIGVATVMEHNLVHSSIVQTRAYP-RTP 272

Query: 250 SSIKKAVRPMTLDNSIQSLAFVFNVSILLVLGASLFG-HANDISAFQMYLALSDKIT 308
 +A+R+ DS + L LA +N+ +L+L A++F+ R D+ Q Y L+
 Sbjct: 273 VGRSALRVAVADSTLALMLALFINASILILAAVVFHQAHRFVSEIEQAYQLAPVLGV 332

Query: 309 GAVASSFLSTLFAVALLASQNSTITGTLTGQIVMBSFLPKQLPQLRLCRLTLTLPLPI 368
 G A TLFA ALLASG NST+T TL GQIVMBSFL +L WL R+ TR L ++P+
 25 Sbjct: 333 GVAA-----TLFATALLASGINSTVATLTAQIVMBSFLRLRLPRLRVLRVGLAIVFV 387

Query: 369 FVIALWGGSENTLQGLIVYSQVFLSALPFSIFPLIYFTSQKSIMGHEHNAKWTYLA 428
 V+ L G E +L++ SQV Ls+ LFF+ EL+ + +M3 +W +A+
 30 Sbjct: 388 IVVVALYG--EQGTGRLLLSQVLSMQLFPAVITLRCVADRNVAGALVAFRWLVVW 445

Query: 429 IWAIIILTLNKLIMD 444
 L+A ++ +LN+KL+ D
 35 Sbjct: 446 LIAGVIVVLNKLGD 461

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 336

40 A DNA sequence (GBSx0366) was identified in *S.agalactiae* <SEQ ID 1087> which encodes the amino acid sequence <SEQ ID 1088>. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have a cleavable N-term signal seq.

45 INTEGRAL Likelihood =-14.12 Transmembrane 113 - 129 (98 - 132)
 INTEGRAL Likelihood =-12.15 Transmembrane 228 - 244 (220 - 249)
 INTEGRAL Likelihood =-10.83 Transmembrane 175 - 191 (167 - 195)
 INTEGRAL Likelihood = -5.04 Transmembrane 57 - 73 (55 - 75)
 INTEGRAL Likelihood = -3.93 Transmembrane 146 - 162 (142 - 166)
 50 INTEGRAL Likelihood = -1.38 Transmembrane 199 - 215 (199 - 215)
 INTEGRAL Likelihood = -0.32 Transmembrane 82 - 98 (82 - 98)

----- Final Results -----
 bacterial membrane --- Certainty=0.6647(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 55 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF11325 GB:AB002018 hypothetical protein [Deinococcus radiodurans]
 Identities = 63/215 (29%), Positives = 108/215 (49%), Gaps = 13/215 (6%)

60 Query: 11 LLIVFILTIVNYLSATGFLTGNQKSLSDRYQTLLTPAPLAFSIVSVIYL-LTFVLILR 69

-426-

```

      LL +LT++VNYLS L GNS +SDR TPA L F++W I+L L + +
Sbjct: 10 LLAATVLT/LVVNYLSNALPLPONSNAEVSDDLPHAPTPAGLTPTVWGPIPLGLLVFAVQ 69

Query: 70 AIPSKSQSYQDNFASIPFPYPLGLLVNNIWTVPFTSNLIGLSTIIIFAYCILLV-IIIKI 128
      A+ ++ + D +P+ LG LL N W + F S IGLS +I+ A +LV + + +
Sbjct: 70 ALPAQRGARLDRL--FWFPLGKLL-NVAMLLAPQSLNIGLSVIMALLAVLVRILYSV 126

Query: 129 LS---KNSKLLRLITPGIHAGMLVASLVMALVYLKI----DFNYPLPKVYIAIALI 181
      S + + L++ ++ W+ VA++ N+ +LV F V+ A++ ++
Sbjct: 127 RSLFPQGAERWTLOLPVSLYLAWISVATITANTAFVLSAGVTSQSLGAGFVMSALLLV 186

Query: 182 FITVLSLYLARVLONAYLILSVFWKMLVFFAHLE 216
      + +L R A+ + + WA+ V+ A E
Sbjct: 187 AALGVPFLNRFRDTAFANV-LLMAFYGVYVARPE 220

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 337

- 20 A DNA sequence (GBSx0367) was identified in *S.agalactiae* <SEQ ID 1089> which encodes the amino acid sequence <SEQ ID 1090>. Analysis of this protein sequence reveals the following:

```

Possible site: 36
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3401(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 The protein has homology with the following sequences in the GENPEPT database:
      >GP:AAC65352 GB:AB001215 T. pallidum predicted coding region
      TP0352 [Treponema pallidum]
      Identities = 28/64 (43%), Positives = 41/64 (63%)

35 Query: 3 EPTFEIVEKLLVLSENEKGWTKELNKRVSFNGAPAKFDLRTWSPDHTKMGKGITLSNEEFK 62
      +F +E+ LS + GW+ EL +S+NG P K+D+R WSPD +KMGK+TL+ E
Sbjct: 12 DFHYEVTRNWATLSTSGNGSLKLSISWNGRPEKYDIRAWSPOKSFMGKGVITLRAIV 71

Query: 63 VILD 66
      + D
40 Sbjct: 72 ALRD 75

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1091> which encodes the amino acid sequence <SEQ ID 1092>. Analysis of this protein sequence reveals the following:

```

45 Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4021(Affirmative) < succ>
50      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

      Identities = 59/70 (84%), Positives = 64/70 (91%)

55 Query: 1 MSEFTFEIVEKLLVLSENEKGWTKELNKRVSFNGAPAKFDLRTWSPDHTKMGKGITLSNEK 60
      M+EFTF I E LL LSEN+GWTKELNKRVSFNGA AK+D+RTWSPDHTKMGKITL+NEE
Sbjct: 1 MAEFTFTEIEHLLTISENKGWTKELNKRVSFNGARAKWDIRTWSPDHTKMGKITLNEE 60

```

Query: 61 FKVILDAPRK 70
 FK IILDAPRK
 Sbjct: 61 FKVILDAPRK 70

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 338

A DNA sequence (GBSx0368) was identified in *S.galactiae* <SEQ ID 1093> which encodes the amino acid sequence <SEQ ID 1094>. Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.66 Transmembrane 92 - 108 (92 - 110)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2062(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB14675 GB:Z99117 similar to protease [Bacillus subtilis]
 Identities = 201/407 (49%), Positives = 277/407 (67%), Gaps = 2/407 (0%)
 Query: 4 VKGRPEVLSPAGTLEKLVAIDYGADAVFVGQAYGLSRAGNPSMELQSGINIAHARD 63
 + K+FE+L+FAG LEKLAA+ YGADAVF+GQ YGLRS A NF++B+ BG+ +A
 Sbjct: 18 ITKKPELLAPAGNLEKLKIAVEYGADAVFVGQYGLSRNADNPTIEIAGVEPAKKYG 77
 Query: 64 AKYVVAAMVTHENELGAGPWFRELKRMGLDAVTSOPALIVICATEAPGLIEHLSTQA 123
 AK+IV N+ H N G + + L D + +IV+DP +I C AP +E+HLSTQ
 Sbjct: 78 AKIYVTINI FAHNEWMDGLBDYKALGADNAVAGIIVADPLIETCRVAENVEHLSTQQ 137
 Query: 124 SSTNYETPFWKEMGLTVVLAREVTMAELAEIRKTDVEIAFVHGAMCISYSGRCVLS 183
 S +N++ +FWKE QL RVVLARE + E+ E+++ D+EIE+P+HGAMCI+YSGRCVLS
 Sbjct: 138 SLENKAVQFWKEGLDRVVLARETSALSTIREMKEKVDIEISPIHGAMCIAYSGRCVLS 197
 Query: 184 NIMSHRDANRGGCQSCRWKYDLYMPFQGERQSLKEIPEPPSMSAVDMCMIEHIPQMI 243
 NIM+ RD+NRGGC QSCRW YDLY G +L GE FP+MG D+ +IE IP MI
 Sbjct: 198 NIMTARDSNRGGCCQSCRWYDLYQTD-GANAVALYGEEDAPFMSPDKLKIESIPQMI 256
 Query: 244 ENGVDLSLKIEGRMKSIHYVSTVTMCYKAAVDAYMESPEAFEAIKDLIDELMKVAQRELA 303
 E G+DSLKIEGRMKSIHYV+TV + Y+ +DAY PE F I+++ +EL K A R+ A
 Sbjct: 257 ENGIDLSLKIEGRMKSIHYVATVVSVYRKVIDAYCADPENF-VIQKNEELDKCANRDTA 315
 Query: 304 TGFVYHTPTENQLPQARRKIPOYKPVGEVVSFINAKMEATIRQNVIMEGRDVEFYVGG 363
 T F+ TP PQ-FG K Y FVG V++D T++QW +GD VEF+GP
 Sbjct: 316 TAFVEGTIPGYEQMFGHAHAKTYDVPGLVLYNVEDTQWVTLQQRNFKFGKDEVEFPGE 375
 Query: 364 PRHFECFDGLDRAEGNKIDRAPNMEILLITLNPVKKGMIRACK 410
 +F I+ + D +GN+D A +P+++ L + +M+R K
 Sbjct: 376 IENFTHTIETINDEGNELDAARHPLQIVKFKLKKIYPSNMKKK 422

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1095> which encodes the amino acid sequence <SEQ ID 1096>. Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.66 Transmembrane 92 - 108 (92 - 110)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2062(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

5 >GP:BAB04993 GB:AP001511 protease [Bacillus halodurans]
Identities = 201/403 (49%), Positives = 280/403 (68%), Gaps = 4/403 (0%)

Query: 6 KRPEVLSPAGTLEKLVKVAIDYGADAVFGGQAYGLRSRAGNFSMEELQSGIDYAHARGAK 65
K+PE+L+PAG+LEKLVKVAI YGADAV++GGQ +GLRS A NFS+EE++EG++A+ GAK
10 Sbjet: 17 KKFPELLAPAGSLKLVKVAIHYGADAVYIGGQFGLRSNAINFSIEMREGEVFAFNKYGAK 76

Query: 66 VYVAANMVTHEGNEIGAGWFRLRDMGLDAVIVSDPALIVICSTRAPGLEIHLSTQASS 125
VYV N+ H N G E+ L+++G+ +IV+DP +I C AP +E+HLSTQ S
10 Sbjet: 77 VYVTINIAHNNMDGLELYLSALQEVGTGIIVADPLIETCKRVAPKVEVHLSTQCSL 136

Query: 126 TNYETFEFWKMGILTRVVLAREVNMAELAEIRKRTDVIEIRAFVHGAMCISYSGRCVLSNH 185
+N+ +FWK GL RVVLAREV +E E+K D+EIE FVHGAMCISYSGRCVLSNH
15 Sbjet: 137 SNWLAVFWKEBGLHRVVLAREVGLSEEMLEMKGVDIETIEFVHGAMCISYSGRCVLSNH 196

Query: 186 MSRDNRNGGCSQSCRWKYDLTYMPFGQE-RRLSLKEI PEDYSMSVDMCMIDHIDPLIE 244
N+ RD+NRGGC QSCRW YDLY+ E +G++P Y+MS D+ +I IP LIE
20 Sbjet: 197 NTRDNRNGGCSQSCRWYDLYTQCSSEIPLFABGDPV--YTNSPDLNLIQAIPQLIE 254

Query: 245 NGVDSLKIEGRMKSIHYVSTVINCYYKAAGVYMSPEAFYAIKEBILDELKVAQRELAT 304
G+DSL+EGRMKSIHYV+TVI+ Y+ +AY D+ F IK E ++EL K A R+ A
25 Sbjet: 255 AGIDSLKIEGRMKSIHYVATVTSYRKVIDAYCSDPNP-KIKREWLELEKCANRDEAP 313

Query: 305 GFYTGIFTNEBQLGARKKIPQKYFVGEVAFDASMTATIRQNVIMSGRIEYCGPFG 364
F+ G PT RQ+G K +Y FVG V+ ++ + T+QRN +GD +E +G
30 Sbjet: 314 QFFGFTYTFEYKGIHFKRTKYDFVGLVDYNEKGIIVTLQQRNIHFQGDVEVFFGPEI 373

Query: 365 RIFETVYKDLHDAQCKIDRAPNMELLITSLPREVKPGDMIR 407
P Y+ + D DG ++D A +P++++ + +V P +N+R
35 Sbjet: 374 NRFTQVVEKTDWEDGNELDARHPLQIVKFKVDQKYVQNMNR 416

An alignment of the GAS and GBS proteins is shown below:

Identities = 386/427 (90%), Positives = 404/427 (94%)

Query: 1 MSNVKRRPEVLSPAGTLEKLVKVAIDYGADAVFGGQAYGLRSRAGNFSMESLQSGINIAH 60
MS++KRRPEVLSPAGTLEKLVKVAIDYGADAVFGGQAYGLRSRAGNFSMEELQSGI+YAH
40 Sbjet: 1 MSNMKRRPEVLSPAGTLEKLVKVAIDYGADAVFGGQAYGLRSRAGNFSMEELQSGIDYAH 60

Query: 61 ARDAKVVYVAANMVTHEGNEIGAGWFLRDMGLDAVIVSDPALIVICSTRAPGLEIHLIS 120
AR AKVYVAANMVTHEGNE+GAG WFLRDMGLDAVIVSDPALIVIC+TRAPGLEIHLIS
40 Sbjet: 61 ARGAKVVYVAANMVTHEGNEIGAGWFLRDMGLDAVIVSDPALIVICSTRAPGLEIHLIS 120

Query: 121 TQASSTNYETFEFWKMGILTRVVLAREVNMALAEIRKRTDVIEIRAFVHGAMCISYSGRC 180
TQASSTNYETFEFWK MSLTRVVLAREV MELAELIRKRTDVIEIRAFVHGAMCISYSGRC
45 Sbjet: 121 TQASSTNYETFEFWKMGILTRVVLAREVNMALAEIRKRTDVIEIRAFVHGAMCISYSGRC 180

Query: 181 VLSNHSRHDNRNGGCSQSCRWKYDLTYMPFGQERGLKEIPEPFSMSAVLMCMIEHIP 240
VLSNHSRHDNRNGGCSQSCRWKYDLTYMPFG ER+SLKGEIPE +SMS+VLMCMI+HIP
50 Sbjet: 181 VLSNHSRHDNRNGGCSQSCRWKYDLTYMPFGGERGLKGEIPEPYSMSAVLMCMIDHIP 240

Query: 241 DLIEKNVDSLKIEGRMKSIHYVSTVINCYYKAAGVYMSPEAFYAIKEBILDELKVAQR 300
D+IENKVDLSLKIEGRMKSIHYVSTVINCYYKAAGV YMSPEAF ATK+L+IELKVAQR
55 Sbjet: 241 DLIEKNVDSLKIEGRMKSIHYVSTVINCYYKAAGVYMSPEAFYAIKEBILDELKVAQR 300

Query: 301 ELATGFTYHTFTNEBQLGARKKIPQKYFVGEVAFDASMTATIRQNVIMSGRIEYCGPFG 360
ELATGFTY FTHNEBQLGARKKIPQKYFVGEVAFD+ A M ATIRQNVIMSGR+3 Y
60 Sbjet: 301 ELATGFTYHTFTNEBQLGARKKIPQKYFVGEVAFDASMTATIRQNVIMSGRIEYCGPFG 360

Query: 361 GPGFHFPECFDILDRADNMCKIDRAPNMELLITSLPREVKPGDMIRACKBGLVNLQYD 420
GPGFHFPE + L DA+G KIDRAPNMELLIT+L D VK GDMIRACKBGLVNLQY D
60 Sbjet: 361 GPGFHFPEVVKDLDADSGKIDRAPNMELLITSLPREVKPGDMIRACKBGLVNLQYD 420

Query: 421 QTSKTVR 427

GTSKTVR

Sbjct: 421 GTSKTVR 427

SEQ ID 1094 (GBS385) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 69 (lane 3; MW 50kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 7; MW 75.7kDa).

The GBS385-GST fusion product was purified (Figure 213, lane 7) and used to immunise mice. The resulting antiserum was used for FACS (Figure 312), which confirmed that the protein is immunoinaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 339

A DNA sequence (GBSx0369) was identified in *S. agalactiae* <SEQ ID 1097> which encodes the amino acid sequence <SEQ ID 1098>. This protein is predicted to be collagenase. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2208 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB14677 GB:Z99117 similar to protease [Bacillus subtilis]
Identities = 92/304 (30%), Positives = 161/304 (52%), Gaps = 5/304 (1%)

Query: 1 MEKILITATARSIEQVKQLLAIGIDRIYVGEENYGLRLPHSFSDDLRLAKLVHDGKE 60
      M+K L T S + L+ G VGE+ YGLRL FS +++ + H G +
Sbjct: 1 MKKPELLVPTTSTADILFLIQAGATAFLVGEQRYGLRLAGEFSREDVTKAVRIKHEGAK 60

Query: 61 LTVACNALMHQBMMDNIKPFLLMLMKEINVDYLVGDAGVFFINKRGDYNFKLYDTSVFV 120
      + VA NA+ H + + + L+ E VD V GD V + + KL+ T
Sbjct: 61 VYVA/NAIHNDKVGLGKYLFLAELAGVDAAVFGDPAVLMARESADFLKLHWSTETG 120

Query: 121 TSSEQVNFVWGCHGAVETVLAREIPSEKLFMSSENLEFPARTLVYGASVIHHSKRPLLQNY 180
      T+ N+WG+ GA +VLARE+ + + + EN E EI V+G + + SKR L+ NY
Sbjct: 121 TNYTTCNYWGRKGAARSVLARELNMDSDIVEIKENAEVRIQVHGMTOMFQSKRLIGNY 180

Query: 181 YNF---THITDEKTRRERGLFLAREPGDPESHYSIYEDKHGTHIFINNDINMVKVTELVEH 237
      + + + K +E G+FL + + + Y I+ED++GTHI ND+ + + + EL++
Sbjct: 181 FEYQCKVMDIERKKKSGMFLHDK-ERDNKPIFEDENGTHTMSNDVCIIDELSELIDA 239

Query: 238 HPTHKWLXGIYCPGDNFVALAEIFVETARI- IENGTFITDQAFLEDERIKLHFKRGGLD 296
      +K+DG+ + + + + E L+EN + + + ERI + P R +D
Sbjct: 240 GIDSKFLDGKMPYLLIEVTMYREAIIDLCVENRDEYBAKKEWIERIESIQFVNKID 299

Query: 297 TGFY 300
      TGF+
Sbjct: 300 TGFF 303
```

A related GBS nucleic acid sequence <SEQ ID 10949> which encodes amino acid sequence <SEQ ID 10950> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1099> which encodes the amino acid sequence <SEQ ID 1100>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1716 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 245/308 (79%), Positives = 273/308 (88%)

Query: 1 MEKIIITATAESIEQVKQLLAIGIDRIYVGEENYGLRLPHSPSDDLEIAKLVDHAGKE 60
MEKII+TATAESIEQVK LLA G+DRIYVGE NYGLRLPH+PS DELR+IAKLVDHAGKE
Sbjct: 1 MEKIIITATAESIEQVKALLAAGVDRIYVGEANYGLRLPHNPSYDELRQIAKLVDHAGKE 60

Query: 61 LTVACNALMHQEMMNIKPFLELMKEINVDYLVGDAGVFFYINKRDGYNFKLIYDTSVFV 120
LTVACNALMHQ+MMD IKPFL+LM EI VDYLVGDAGVFFY+NKRDGYNFKLIYDTSVFV
Sbjct: 61 LTVACNALMHQDMMQIKPFLDLMIETAVDYLVDGAGVFFYNNKRDGYNFKLIYDTSVFV 120

Query: 121 TSSRQVNFQGHGAVETVLAREIPSEELFKMSNLEFPASILLVYGASVIHHSKRPLQNY 180
TSSRQVNFQGHGAVE+VLAREIPS ELP ++ENLEFPAB+LVYGASVIHHSKRPL+NY
Sbjct: 121 TSSRQVNFQGHGAVESVLAREIPSAELFTLAENLKFPAVSVLVYGASVIHHSKRPLLENY 180

Query: 181 YNFTHITDEKTRERGLFLAEPGDPSHSYIYEDHKGTHIFINNDINMMIKVTLVBHHT 240
Y+FT I DE +RERGLFLAEPGD SHSYIYED HGTHIFINNDI+MM+K+ RL H T
Sbjct: 181 YHFTKIDDEVSERERGLFLAEPGDASSHSYIYEDHNGHTHIFINNDIDMMKSLERLVAHGILT 240

Query: 241 HWKLDGIYCPGDDFVAIAEIVFETARLIENGFTYDQAPFLDERIKLHPKRGLLITGYF 300
HWKLDGIYCPGD+FVAI ++F++ L+E G FTQ+AA D+ + HP GRGLDTGYF
Sbjct: 241 HWKLDGIYCPGDDFVAITKLFQAKTLLEAGQPTQREARKLDQAVHAHHPAGRGLOTGYF 300

Query: 301 FDPDPSTVK 308
+FDP TVK
Sbjct: 301 EFDPTVK 308
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 340

A DNA sequence (GBSx0371) was identified in *S.agalactiae* <SEQ ID 1101> which encodes the amino acid sequence <SEQ ID 1102>. This protein is predicted to be cDNA EST yk542c12.5 comes from this gene. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD15622 GB:U75480 unknown [Streptococcus mutans]
Identities = 69/152 (45%), Positives = 101/152 (66%), Gaps = 12/152 (7%)
```

```
Query: 1 MSKLPKTLVISAASGAAAAYFLITKKGKELRQNAEKFYGRYKENPBEYTHQIAKKDASEYS 60
MSK KT +I A +GAAAAYFL+T KKK+ +K + +YKENP+EYHQ A DK +EY
Sbjct: 1 MSKPLKTAIIGAGTGAAAAYFLSTDKGQFKKKIHQITFDYKENPKETHQYAADKVNSEYK 60
```


-431-

Query: 61 NLAVDTFQDYKGFESGELITTEDIVSAVKESGEVDFANDFVNQAKSKFSDDETAKEKED 120
 ++AV +PKDYK KFE+GELT ++I+S+VKKK+ + FAN ++Q K + T +K +
 Sbjct: 61 DVAVHSFDYDKRFPETGBLITRENIISVVKESQAGKAFNSKLSQVKKHLA--QTVKES 118

Query: 121 KAP-----ETKVEDIVIDYKENTSDKE 142
 + +V+DIVIDY+ +K+
 Sbjct: 119 ASTNDAGIPLGEMKQVDDIVIDYQASEKTK 150

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1103> which encodes the amino acid sequence <SEQ ID 1104>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.81 Transmembrane 15 - 31 (14 - 31)

----- Final Results -----
 bacterial membrane --- Certainty=0.1723 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9117> which encodes the amino acid sequence <SEQ ID 9118>. Analysis of this protein sequence reveals the following:

Possible cleavage site: 19
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty= 0.300 (Affirmative) < succ>
 bacterial membrane --- Certainty= 0.000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty= 0.000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 69/140 (49%), Positives = 91/140 (64%), Gaps = 3/140 (5%)

Query: 1 MSKLFKTLWISAASGAAAYFLTTKKGKELRQNAEKFYGEYKENPPEYHQIAKDKASEYS 60
 M+K FK LVI A SG AAAYFL+T+XGK L+ AEK Y YKE+P++YHQ AK+K SEYS
 Sbjct: 8 MNKSFQILVIGAGSVAAAYFLSTERGKALNRAEKAYQAYKESPDYHQFAKEKSEYS 67

Query: 61 NLAVDTFQDYKGFESGELITTEDIVSAVKESGEVDFANDFVNQAKSKFSD-EDTAKE 119
 ++LA DTF D K K SG+LT ED++ ++K+K+ FV + K ++ E K++
 Sbjct: 68 HLRDTTFYDVKDKLASGLTKEDMLDLKDKT-----TAFVQKTKSLASVEAKEKQD 120

Query: 120 DKAPETKVEDIVIDYKENTE 139
 D + EDI+IDY E E
 Sbjct: 121 DVIIDLNREEDIIIDYEQDE 140

SEQ ID 1102 (GBS164) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 30 (lane 4; MW 17.4kDa).

The GBS164-His fusion product was purified (Figure 115A; see also Figure 200, lane 4) and used to immunise mice (lane 1+2+3 product; 20µg/mouse). The resulting antiserum was used for Western blot, FACS (Figure 115B), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoreactive on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 341

A DNA sequence (GBSx0372) was identified in *S.agalactiae* <SEQ ID 1105> which encodes the amino acid sequence <SEQ ID 1106>. Analysis of this protein sequence reveals the following:

```

Possible site: 19
>>> Seems to have an uncleavable N-term signal seq
      INTEGRAL      Likelihood = -16.93      Transmembrane      6 - 22 ( 1 - 31)

----- Final Results -----
      bacterial membrane --- Certainty=0.7771(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD15621 GB:U75480 unknown [Streptococcus mutans]
Identities = 88/129 (68%), Positives = 112/129 (86%)

Query: 1 MIEIAVLIIAIAFVVLVLGLFVLKKVSEITRETKQTIKVLTSDEVNVTLYQTNELAKAN 60
      N EIA+LI+AIAP VLV+ ++ +L+K+S+T++E++QT+K+LTSDEVNVTLYQTNELAKAN
Sbjct: 1 MWEIALLIVAIAFAVLVITLYLLLRKISDTVDESQTLKILTSDVNVTLYQTNELAKAN 60

Query: 61 VLVDVNGKVSITIDPLFVAIADLSSEVSDNLNQAHHQKQKASATSSVTGASALAIKFA 120
      VLV+DVNGKV TIDPLF AIADLS SVSDLN QAR+ G+K +T++V KAG+A GK
Sbjct: 61 VLVDVNGKVETIDPLFTAADLSVSVSDLNQARLYFGKKTRKSTANVGKGAAYTFGKV 120

Query: 121 ASKIFPKKG 129
      ASK+PKKG
Sbjct: 121 ASKLFPKKG 129

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1107> which encodes the amino acid sequence <SEQ ID 1108>. Analysis of this protein sequence reveals the following:

```

Possible site: 16
>>> Seems to have a cleavable N-term signal seq.
      INTEGRAL      Likelihood = -0.85      Transmembrane      18 - 34 ( 17 - 34)

----- Final Results -----
      bacterial membrane --- Certainty=0.1341(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAD15621 GB:U75480 unknown [Streptococcus mutans]
Identities = 83/128 (64%), Positives = 110/128 (85%)

Query: 6 ISLMIIAIAFVAIVFLIIVLAKVSEITDEAKKTSIVLTSDEVNVTLYQTNELAKANILV 65
      I+L+I+A+AF LVI+LI+L+K+S+T+DE+++T+ +LTSDEVNVTLYQTNELAKAN+LV
Sbjct: 4 IALLIVAIAFAVLVITLYLLLRKISDTVDRSQTLKILTSDVNVTLYQTNELAKANILV 63

Query: 66 EDVNGKVATIDPLFVAIADLSLSLSDLNQAHHQKQKATNATGNVSKAGKIALVNGVASK 125
      EDVNGKV TIDPLF AIADLS S+SDLN QAR+FG+K +T NV KAG GKVASK
Sbjct: 64 EDVNGKVETIDPLFTAADLSVSVSDLNQARQRYFGKKTRKSTANVGKGAAYTFGKVASK 123

Query: 126 VFGKKGEK 133
      +F KKG++
Sbjct: 124 LFRKKGKQ 131

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 92/131 (70%), Positives = 116/131 (88%)

Query: 1 MIEIAVLIIAIAFVVLVLGLFVLKKVSEITRETKQTIKVLTSDEVNVTLYQTNELAKAN 60
      ++ I+++IIA+AFV LV+ ++ VLKKVSEIT+E K+TI VLTSDVNVTLYQTNELAKAN

```

-433-

Sbjct: 3 LVGISLMIIALAFVALVIFLITVLKKVSETIDEAKKTSVLTSDVNVTLHQNDILAKAN 62

Query: 61 VLVDVNGKSVTIDPLFVALADLSSEVSIDIALQARHIGQKASSATSSVTGAOSALAIGKA 120
+LV+DVNGKV+TIDPLFVALADLSBS+SDLN QARH GQKA++AT +V+KAG +GK

5 Sbjct: 63 ILVEDVNGKVATIDPLFVALADLSSELSIDLASQARHFGQKATNATGNVSRAGLALVGKV 122

Query: 121 ASKIFRKKGDK 131
ASK+P KKG+K

10 Sbjct: 123 ASKVPGGKGGK 133

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 342

A DNA sequence (GBSx0373) was identified in *S. agalactiae* <SEQ ID 1109> which encodes the amino acid sequence <SEQ ID 1110>. Analysis of this protein sequence reveals the following:

Possible site: 13
>>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----

bacterial cytoplasm	---	Certainty=0.0462 (Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 343

A DNA sequence (GBSx0374) was identified in *S. agalactiae* <SEQ ID 1111> which encodes the amino acid sequence <SEQ ID 1112>. This protein is predicted to be prolipoprotein diacylglycerol transferase (lgt). Analysis of this protein sequence reveals the following:

Possible site: 29
>>> Seems to have an uncleavable N-term signal seq

35 INTEGRAL Likelihood = -8.39 Transmembrane 231 - 247 (225 - 251)
INTEGRAL Likelihood = -7.64 Transmembrane 89 - 105 (87 - 107)
INTEGRAL Likelihood = -5.20 Transmembrane 18 - 34 (13 - 36)
INTEGRAL Likelihood = -1.06 Transmembrane 46 - 62 (46 - 64)

40 ----- Final Results -----

bacterial membrane	---	Certainty=0.4354 (Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

A related GBS nucleic acid sequence <SEQ ID 9721> which encodes amino acid sequence <SEQ ID 9722> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC80171 GB:U75480 putative prolipoprotein diacylglycerol
transferase [Streptococcus mutans] (ver 3)
Identities = 184/257 (71%), Positives = 226/257 (87%)

50 Query: 2 MINPVAIRLGPFSSIRNYAICIVSGMLLAVYLAMKEAPRNKISDDILDILFIAMAFPLSIGV 61
MINP+AI+LGP +IRNY+ICIV+G+LAVYL ++EAP+KNISDD+LDLFI+AFPL+IVG

-434-

5
 10
 15
 20
 25
 30
 35
 40
 45
 50
 55
 60

```

Sbjct: 1  MINPIAIKIGPLTIRWYSICITVGLILAVYLITIRAPKKNKSDVDLFDILIAFPFLAIVG 60

Query: 62  ARYYVVFPMWYYSKHPVELIATWNGGIAIYGGILITGAILLVIFSYRRLINPFDIDIAA 121
          AR+YVVF+M YY K+P EI  IW+GGIAIYGGIL+TGAA+L IFSY R+I PIDFID+AA
Sbjct: 61  ARLYYVIFPMWYYSKHPVELIATWNGGIAIYGGILITGAILLVIFSYRRLINPFDIDLVAA 120

Query: 122  PGVMTAQAIQKGNFNPINQKAYGKAVKNLNVNPFKQKMYIDGAYRVPTFLYESLWNLG 181
          PGVM+AQ+IGKGNF+NQKAYG+ V  LAY+P+PI+ QNYIDG YR PTFLYESLWNLG
Sbjct: 121  PGVMTAQSIGKGNFNPVQKAYGKTVIQLNLYLDFIRKQMYIDGHYKPTFLYESLWNLG 180

Query: 182  FVIIMSRHPRPTLKQGEVACFYLIWYGGGRFIIIGMRTDGLYAGLVSQWLSVLVII 241
          F+IIM +R RP  LK+GEVA FYL+WYG GRF+IIGMRTDGL A LRVSQWLSV+IV++
Sbjct: 181  FIILMILRRRNLKQGEVAFYLIWYGGGRFVIEGMRTDGLMFASLRVSQWLSVLVVV 240

Query: 242  GIVMIIYRRERQHSYY 258
          G++++ RRR  I YY
Sbjct: 241  GVILMVIRRRNHAIPIYY 257
  
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1113> which encodes the amino acid sequence <SEQ ID 1114>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -7.01 Transmembrane 229 - 245 (222 - 249)
 INTEGRAL Likelihood = -6.90 Transmembrane 45 - 61 (40 - 68)
 INTEGRAL Likelihood = -4.41 Transmembrane 17 - 33 (11 - 35)
 INTEGRAL Likelihood = -4.14 Transmembrane 87 - 103 (86 - 106)
 INTEGRAL Likelihood = -0.27 Transmembrane 170 - 186 (170 - 186)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.3803 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

35
 40
 45
 50
 55
 60

```

>GP:AA08171 GB:U75480 putative prolipoprotein diacylglycerol
transferase [Streptococcus mutans] (vex 3)
Identities = 176/258 (68%), Positives = 217/258 (83%)

Query: 1  MINPIALKOPLATHRYALCTILSGILAVLYLASKRAPKGISSDAIFDILIAFPFLAIVG 60
Sbjct: 1  MINPIA+K GEL I KY++CI++GL+LAVYL +EAPK I SD +DFILIAFPFLAIVG 60

Query: 61  ARYYVVFPMWYYSKHPVELIATWNGGIAIYGGILITGAILLVIFSYRRLINPFDIDIAA 120
          AR+YVVF+M YY+K+ EI  IW+GGIAIYGGIL+TGAILV +Y +++ PI FLD+AA
Sbjct: 61  ARLYYVIFPMWYYSKHPVELIATWNGGIAIYGGILITGAILLVIFSYRRLINPFDIDLVAA 120

Query: 121  PGVMTAQAIQKGNFNPINQKAYGKAVSNLYLPSPIQKQMFEGSYRPTFLYESLWNLG 180
          P VM+AQ+IGKGNF+NQKAYGK V+QMYLDF PI+KQ+I+G YR PTFLYESLWNLG
Sbjct: 121  PGVMTAQSIGKGNFNPVQKAYGKTVIQLNLYLDFIRKQMYIDGHYKPTFLYESLWNLG 180

Query: 181  FVIIMSRHPRPTLKQGEVACFYLIWYGGGRFIIIGMRTDGLYAGLVSQWLSVLVII 240
          F+IIM+R R+P L +GE+ FYLIWYGGGR VIEGMRTDGLMF +R+SQ+S LL+++
Sbjct: 181  FIILMILRRRNLKQGEVAFYLIWYGGGRFVIEGMRTDGLMFASLRVSQWLSVLVVV 240

Query: 241  GLIFVIKRRERQHSYYQ 258
          G+I ++ RRR  I YYQ
Sbjct: 241  GVILMVIRRRNHAIPIYY 258
  
```

An alignment of the GAS and GBS proteins is shown below:

60

```

Identities = 176/257 (68%), Positives = 221/257 (85%)

Query: 2  MINPVAIRLGPFSSIMYALCIVSGMLLAVLYLAKKAPKIKSDSDIIFDILIAFPFLAIVG 61
          MINP+AAA GP +I WYA+CI+SG++LAVLYLAKKAP+K I SD I DFIL+APFL+IVG
Sbjct: 1  MINPIALKOPLATHRYALCTILSGILAVLYLASKRAPKGISSDAIFDILIAFPFLAIVG 60
  
```

```

5 Query: 62 ARIYYVIFEMAYYSKPIPELAIINMGSGIALYGGGLITGAILVIFPSYRRLINPIDFLDIA 121
   Sbjct: 61 ARIYYVIFEMAYYKH KETAIINMGSGIALYGGGLITGA+L+ + Y + +NPI FLIDIA
   Query: 122 PGVMIAQAIGRWGNFINQYKAYGRVKNLYVPNFINKQMYIDGAYRVPTFLYBSLANITLG 181
   Sbjct: 121 P VMIAQAIGRWGNFINQYKAY+LNY+L+Y+ QM+L+G+YR+PTFLYBSLAN LG
   10 Query: 182 FVIMISIRHRPRTLKGEVAFILVNWYGGGRFIIEMGMTDSLYLAGLRVSONLSVLVII 241
   Sbjct: 181 FVILN R +P+L+L GR+ PVL+WG GR +LBMGTMSL G+R+SQ+L+L+I
   Query: 242 GIVMIYRRREQHISY 258
   15 Sbjct: 241 GLIPVKKRRQKGISY 257

```

A related GBS gene <SEQ ID 8557> and protein <SEQ ID 8558> were also identified. Analysis of this protein sequence reveals the following:

```

20 Lipop: Possible site: -1 Crend: 0
   McG: Discrim Score: 2.45
   GVH: Signal Score (-7.5): -2.9
      Possible site: 39
   >>> Seems to have an uncleavable N-term signal seq
25 ALOM program count: 3 value: -8.39 threshold: 0.0
   INTEGRAL Likelihood = -8.39 Transmembrane 209 - 225 ( 203 - 229)
   INTEGRAL Likelihood = -7.64 Transmembrane 67 - 83 ( 65 - 85)
   INTEGRAL Likelihood = -1.86 Transmembrane 24 - 40 ( 24 - 42)
   PERIPHERAL Likelihood = 0.79 92
30 modified ALOM score: 2.18

*** Reasoning Step: 3

----- Final Results -----
35 bacterial membrane --- Certainty=0.4354(Affirmative) < success
   bacterial outside --- Certainty=0.0000(Not Clear) < success
   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < success

```

The protein has homology with the following sequences in the databases:

[illegible]

```

918      948      978      1008      1038      1068      1098      1128
RTDSL YLAGLRVSGMLSVILVIIGIVMIYRRRBQHSYY*TEEVL**KLLY*LLPLRLLP*F*EYFSP*KKYQKRLRKP
5  ||||| :| ||||| ||||| :||:::|:::| :| :| :|
RTDSL MFA SLRVSGMLSVLLVVGVI LNVIRRRNHAI PYYQC
      230      240      250

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 344

- 10 A DNA sequence (GBSx0375) was identified in *S.agalactiae* <SEQ ID 1115> which encodes the amino acid sequence <SEQ ID 1116>. Analysis of this protein sequence reveals the following:

```

Possible site: 31
>>> Seems to have no N-terminal signal sequence

```

- 15 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2817(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 The protein has homology with the following sequences in the GENPEPT database:

```

>GP:EAA77782 GB:AB027460 Hpr kinase [Streptococcus bovis]
Identities = 264/309 (85%), Positives = 292/309 (94%)

```

- 25 Query: 1 MAVTVQMLVDRKLNVIYGDHLLSKRITTADISRGLEMTGYFDYYAPERLQLVGMKEW 60
M+VTY+MLVD+KL+VIYD+ LLSK ITT+DISRGLEMTGYFDYY+PERLQL+GMKEW
Sbjct: 1 MSVTVMQLVDKVLVDVIYGDHLLSKRITTSDISRGLEMTGYFDYSPERLQLLGMKEW 60
- 30 Query: 61 SYLNAMTGNHREYQVLREMPQKETPAIVVARDLEIPEENYEAAKDTGIALIQSKAPTSLRS 120
SYL MT HNR VLREM +ETPAI+VAR+L IPEEN AAK+GIALIQS PTLRS
Sbjct: 61 SYLTGMVSHNRHVLREMKETPAITVARNLATPEEMISAKEGIALIQSHVPTSLRS 120
- 35 Query: 121 GEVSNYLDSCLAERTSVHGVLMDIYGMGVLIQGSIGKSETGLELVKRGHRLVADDRVD 180
GR+SNYLDSCLAERTSVHGVLMDIYGMGVLIQGSIGKSETGLELVKRGHRLVADDRVD
Sbjct: 121 GEMSNYLDSCLAERTSVHGVLMDIYGMGVLIQGSIGKSETGLELVKRGHRLVADDRVD 180
- 40 Query: 181 VYAKDEETLMGEPAEILRLHLLEIRGVGIIIDMSLYGASAVKDSQVQLAIYLENFETGKV 240
V+AKDEETLMGEPAEILRLHLLEIRGVGIIIDMSLYGASAVKDSQVQLAIYLENF+EGKV
Sbjct: 181 VFADDEETLMGEPAEILRLHLLEIRGVGIIIDMSLYGASAVKDSQVQLAIYLENFESQKV 240
- 45 Query: 241 FDLRGNGNEELSGVTVPRIRIPVKTGRNVSVITEAAMNHRAKQMSFATCTFEDELIT 300
FDLRGNGNEE+ELSGVK+PR+RIPV+TGRN+SVITEAAMNH+RAKQMSFAT+TFE+RLT
Sbjct: 241 FDLRGNGNEELSGVKIPRIPIPVCTGRNMSVITEAAMNYRAKQMSFATKTTFEERLT 300
- Query: 301 HLISGNEVN 309
LI+NE N
Sbjct: 301 QLITSGNEVN 309

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1117> which encodes the amino acid sequence <SEQ ID 1118>. Analysis of this protein sequence reveals the following:

```

Possible site: 13
>>> Seems to have no N-terminal signal sequence

```

- 55 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2391(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 255/309 (82%), Positives = 288/309 (92%)

```

-437-

Query: 1 MAVTVQMLVDRLLKMWIYGDHLLSKRITTADISRPGLEMTGYFDYVAPERLQLVGMKEW 60
 M VTY+GLV ++RL+V+Y ++LLSK ITT+DISRPGLEMTGYFDYVAPERLQL GMKEW
 5 Sbjct: 32 MVTYVQMLVQKVLDDVYATONLLSKRITTADISRPGLEMTGYFDYVAPERLQLVGMKEW 91

Query: 61 SYLMAVMTGHNRYQVIREMPQKETPAIVVARDEIPEEMYTEAAKUTGIALIQSKAPISRLS 120
 SYL MT HRRY VL+EMF+K+TEA+VV+R+L IP+EM +RAK+ GI++L S+ TSLR+
 Sbjct: 92 SYLTQMTSINRYSLKEMFKKUTPAVVVSRNLAIPEKENVQAARKEGILSLSSRVSTSLA 151

10 Query: 121 GEVSVYLDCLARTSVSHVGLMDIYGMGVLIQCDGSGICKSETQLRLVKRCHRIVADRDV 180
 GE+S++LD+ IARSTSVHGLMDIYGMGVLIQCDGSGICKSETQLRLVKRCHRIVADRDV
 Sbjct: 152 GEMSYFLDASLARTSVSHVGLMDIYGMGVLIQCDGSGICKSETQLRLVKRCHRIVADRDV 211

15 Query: 181 VYAKDESTLWGEPAHILRHLLRIBGVGIIDIMSLNGASAVKDSQVQLAIYLENPETGV 240
 VYAKDESTLWGEPAHILRHLLRIBGVGIIDIMSLNGASAVKDSQVQLAIYLENPETGV
 Sbjct: 212 VYAKDESTLWGEPAHILRHLLRIBGVGIIDIMSLNGASAVKDSQVQLAIYLENPETGV 271

20 Query: 241 FDRLGNGNEBIHLSGVKVRIRIPVKTGRNVSVVIEAAMNHRAKMGFDATQTFEDRLT 300
 FDRLGNGNEBIHLSGVKVRIRIPVKTGRNVSVVIEAAMNHRAKMGFDATQTFEDRLT
 Sbjct: 272 FDRLGNGNEBIHLSGVKVRIRIPVKTGRNVSVVIEAAMNHRAKMGFDATQTFEDRLT 331

25 Query: 301 HLISQNEVN 309
 LI++NEV+
 Sbjct: 332 QLITKNEVS 340

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 345

A DNA sequence (GBSx0376) was identified in *S. galactiae* <SEQ ID 1119> which encodes the amino acid sequence <SEQ ID 1120>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> seems to have no N-terminal signal sequence

35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1836 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9719> which encodes amino acid sequence <SEQ ID 9720> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

45 Example 346

A DNA sequence (GBSx0377) was identified in *S. galactiae* <SEQ ID 1121> which encodes the amino acid sequence <SEQ ID 1122>. Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> seems to have an uncleavable N-term signal seq

50 INTEGRAL Likelihood = -4.88 Transmembrane 35 - 51 (31 - 59)

----- Final Results -----
 bacterial membrane --- Certainty=0.2954 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-438-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

- 5 >GP:AAC67275 GB:AP017113 Yv1C [Bacillus subtilis]
Identities = 21/63 (33%), Positives = 36/63 (56%), Gaps = 2/63 (3%)
- Query: 3 SSFYQRKRGKLVGVAGLADKYNDLALSRVLIALILYFTKFP--GLLLYILLAVALFLPYK 60
+ Y+ K K + GV+ GLA+ +ND +L RV+ ++ T LL+YI+ +P +
10 Sbjct: 2 NKLYRSEKNKRIAGVIGGLAEPYFNDASLLRVITVIAIMTSVLPVLLIYIWIPIVPS 61
- Query: 61 EDI 63
D+
Sbjct: 62 RDM 64
- 15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1123> which encodes the amino acid sequence <SEQ ID 1124>. Analysis of this protein sequence reveals the following:
- Possible site: 32
>>> Seems to have an uncleavable N-term signal seq
20 INTEGRAL Likelihood = -5.26 Transmembrane 39 - 55 (31 - 61)
- Final Results -----
bacterial membrane --- Certainty=0.3102 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
25 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

- Identities = 60/90 (66%), Positives = 77/90 (84%), Gaps = 3/90 (3%)
- 30 Query: 1 MKSSFYQRKRGKLVGVAGLADKYNDLALSRVLIALILYFTKFGLLLYILLAVALFLPYK 60
+++ FYKRRK +LV GV+AGLADKY NDAL+RVL AL++Y T FG+LLYILLA+FLPYK
Sbjct: 1 VETKFKQRKNRLVAGVIGGLADKYNDLALSRVLIALILYFTGFGVLLYILLAVALFLPYK 60
- Query: 61 EDIETR-RQGPFRKKDAEPV--DDDGWFW 87
ED++E R +GPRRRKDA+ + +DGWFW
35 Sbjct: 61 EDLLERYGRGPRRRKDAVLEEDDGWFW 90

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 347

- 40 A DNA sequence (GBSx0378) was identified in *S.galactiae* <SEQ ID 1125> which encodes the amino acid sequence <SEQ ID 1126>. Analysis of this protein sequence reveals the following:

- Possible site: 19
>>> Seems to have no N-terminal signal sequence
- 45 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3577 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 50 A related GBS nucleic acid sequence <SEQ ID 9717> which encodes amino acid sequence <SEQ ID 9718> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

- >GP:BAE04250 GB:AP001508 unknown conserved protein [Bacillus halodurans]
Identities = 379/729 (51%), Positives = 515/729 (69%), Gaps = 25/729 (3%)
- 55 Query: 29 ENLNTQIAIDLGIKASQIRKVLRLTDEGNTIPFIARYRKEMTGNLDEVQIKSIDLDEK 88

-439-

```

      E  I  +A +L +K + I++V++L  EGMT+PPTIARYKRK+TG +DEV+I+ + +
Sbjct: 8  EEHTIKTKLAKELSLKNTIKQVQLLHGNTVPTIARYKRKLTGMDGVKIRBVSKEWTY 67

5  Query: 89  MTLSDRKTTVLAKIEBQKGLTCEKLGKIEBATKIADVSELYLPHYENRRKTATIAEAG 148
      L +RK V+ +EBQKGLT E KK +E+A KL +VE+LY PHY+KRR+AT+A+E G
Sbjct: 68  ANGLHRRKEEVRIRLVEBQKGLTDEWKKTVBQKQLQSEVDELFRPHYKRRKTATVAKENG 127

      Query: 149  LFFPLARLI--LQNKDNLESAQNLYLTGPFETT--KALSGAVDILIRAFSEDNKLRSWTY 204
      L FLA + L + +RA+ YL+ E T L GA DI+ E ++D LR
10  Sbjct: 128  LEPLAEWLFLSLPRDGDPLQBAEYVLSVEHELTKVEDVLQGAQDILIEWIADADLRKRIR 187

      Query: 205  NEIMWYSITAVVDSLDRKQVFKIYYDFSEKISKLHGQVLAIRNGENMGVLKVNFEH 264
      + + S+ A VK E LDEK V+++YD+ E + L ++ LALNRGEK VL+V
15  Sbjct: 188  SLGPKSGSVIAKVKKELEBKGVYEMTYDYSEFVKTLVPHRTLALNRGEKEDVLRTVIRF 247

      Query: 265  NLEKMFRRF---FAVRFKETS-QYIDDLIVCTKKKVPAMERRIRTELSBGSBDGLISL 319
      ++++ F RF + Y+ I K+ I P+ER IR EL+E AE+ AI +
20  Sbjct: 248  PVDRILEMSEKTPIRRFSGSPAVPYKAAIEDGQKRLIEPSIEREIRHELEKAGEQAIHI 307

      Query: 320  PSENIENLLVSPKMGVVLGFDPAPRTGAKLAVVDQTKIMTTCVITYPPPNQAKIEQ 379
      F+ENLR+LLL P+KSK+VLG DPA+RTG KLA+VD+TKG++ QVIYD PP N+ +
25  Sbjct: 308  FAENLRSEILLQPPKHKVVLGLDPAKYRTCKLAIVDETGKVLDIQVITYPTPPKNE--VAA 365

      Query: 380  SKIELAKLTKRNFETIAIGNGTASRESAPVAEVLQDPFD-VSYIVNWSGASVYSAGE 438
      +K + KL I ++ +E+IAIGNGTASRES E F+A++++D P + Y+IVN+GASVYSAGE
30  Sbjct: 366  AKKIVKKLIADYGVEMIAIGNGTASRESBQFIADLIKDLQPTIYIIVNWSGASVYSAGE 425

      Query: 439  LARHEFPDLTVKRSASIAARRLQDPLAELVKIDPKSIGVQYQHDVSGKLAEMLPVV 498
      + R EFPDL VB+RSA+SIARRLQDPLAELVKIDPKSIGVQYQHDVSGK+L E+L FVV
35  Sbjct: 426  IGRHEFPDLQVRSASIAARRLQDPLAELVKIDPKSIGVQYQHDVSGKRLNESIFVV 485

      Query: 499  ETVVNQGVGNVNTASPAALLAHVSGLNKTIENIVKYKEENGQIKSRAEIKVPRLGAKF 558
      ETVVNQGVGNVNTASPA+LL +V+GL+KT++NIVK REE G+ +RA++K +ERLGAK +
40  Sbjct: 486  ETVVNQGVGNVNTASPSLLQYVAGLSKTVAKNIVKKEEAGRFTARAQLKIDPRIGAKTY 545

      Query: 559  EQAAGFLRIPNAPKFLPMWIGVHPESYRAVKKLLDGLTIKELD--DLAKKLQNDLILAT 615
      BQ GPRLI + N ID T +HPESY+ KLL ++ D + K+KLQ LD+ A
45  Sbjct: 546  EQCIGFLRIMDGNLNLATAHPESYKVTOKLLESGVATAADVGIEDLKKLQALVSNM 605

      Query: 616  AESIGVGQETLKDIIEDLLKPRDLRDDFEAPVLEHDVLEVSGLKVGQELQGVTRNVVDF 675
      A ++ VQ TLKD+I+ L++P RD RD+ P+L+ DVL + DL G ELQGVTRNVVDF
50  Sbjct: 606  AATLDVGVFTEKMDIALIRPTRDRDEVAKPLKQVLQLEDLLPGMELQGVTRNVVDF 665

      Query: 676  GAFVDIGVHEDGLHQHSLKRRKDKTKRMPPFLKHSKYLGVGDIVTWVWVDAERSR 735
      G FVDVIG +DGL+H S+L R ++HP + ++VG+IVTWVW +VD ++ R
45  Sbjct: 666  GVFDVIGVQDGLVHISKLANKY-----IKHPLVVVGEIVTWVWVWVDDIKKGR 715

      Query: 736  IGLSLIKPD 744
      I L++++P+
50  Sbjct: 716  IALTMLRPE 724

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1127> which encodes the amino acid sequence <SEQ ID 1128>. Analysis of this protein sequence reveals the following:

```

Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2207 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
60  bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 532/716 (74%), Positives = 619/716 (86%), Gaps = 10/716 (1%)
65  Query: 28  MENLNIYQLAINDIGIKASQIEKVLRLTDEGNTIPPTIARYKRKSGMDLDRQKSLIDLEAK 87

```

-440-

MEN N IA L + QIE+VL IT +GNTIPFIARYRKE+TGNLEOV IKSID+DK
 Sbjct: 1 MERNNGHNITABALGVSILHQIEQVIALTAQGNTIPFIARYRKEVTCNLEOVTKSEIIDNDK 60

Query: 88 SMTALSDRKTIVLAKIBEQGKLTQELKAKEATKGLADVELYLFYKSKRRKATATAREA 147
 S+T L+RRK T+LAKIBEQGKLT +L+ +IE KLAD+HEL YLFYKSKRRKATATAREA
 Sbjct: 61 SLTTLNKRKATILAKIBEQGKLTQELTSEATSEKGLADELYLFYKSKRRKATATAREA 120

Query: 148 GLFPLARLLIQNKNDLEEAQNYLIDGPETTTKALSGAVDILIEAFSEDNKLRSWTYNEI 207
 GLFPLARLLIQN NLE A+ ++T+GF + +AL+GAVDIL+EA SED KLRSWTYNEI
 Sbjct: 121 GLFPLARLLIQNAQNLTAASFPVTSASPSQALAGAVDILVEAPMSSEAKLRSWTYNEI 180

Query: 208 VNYSSITAVVQDESIDKQVKFIYDFSEKISKSLAGYVLAALNRGKRMGLKVNFEHLE 267
 W YS + + +KDE LDEK+VF+IYDFS+++S + GY+ LALNRGEK+G+LKV+FEHLE
 Sbjct: 181 WQYSLVSLTKDQDLDEKQVFOIYDFSDQVSNMOGYRTIALNRGKGLKLVSEHLE 240

Query: 268 KMRFFFAVRFKET+SQYIDDLIVOTVKKKIVPAMERRKTELSEGAEDGAILFSENLRNL 327
 KM RFF+VRFKET+ YI+++I QT+KKKIVPAMERR+R+ELS+ AEDGAI LFSENLR+L
 Sbjct: 241 KMORFFSVRFKETNPFYIEVINGTINKKIVPAMERRVSELSDAEDGAILHFSENLRHL 300

Query: 328 LLVSELGKSNVLGDFDAFRTGAKLA+VDQTKGLITQVIYVPPANQAKIBEQSKIELAKL 387
 LLVSELGKSNVLGDFDAFRTGAKLA+VDQTKGL+TYQVIYV PA+Q KI+ +K L +L
 Sbjct: 301 LLVSELGKSNVLGDFDAFRTGAKLA+VDQTKGLITQVIYVAPASQTKIQAARLTLQL 360

Query: 388 IKEFNIIIIAIGNGTASRESEAFVAVLQDFFDVSYIVNESGASVYASSELARHHEFDL 447
 I+ + I+IIAIGNGTASRESEAFV+VL+DFF+ SYIVNESGASVYASSELARHHEFDL
 Sbjct: 361 IETQYDIIIIAIGNGTASRESEAFVADVLDFFNTSYIVNESGASVYASSELARHHEFDL 420

Query: 448 TVEKRSIAISIRRLQDPLAELVKIDPESIGVGQYCHDVSKKLARNLDFVETVNVQGV 507
 TVEKRSIAISIRRLQDPLAELVKIDPESIGVGQYCHDVSKKLARNL FV+VNVNVQGV
 Sbjct: 421 TVEKRSIAISIRRLQDPLAELVKIDPESIGVGQYCHDVSKKLARNLDFVETVNVQGV 480

Query: 508 NVNTASPDALLANVSGLNKTISENIVKYRENGQIKRSABIKVPRLGAKAFQAGFLRI 567
 NVNTASP+LLANVSGLNKTISENIVKYRENG + GRA+IKVPRLGAKAFQAGFLRI
 Sbjct: 481 NVNTASPDALLANVSGLNKTISENIVKYRENGALTSRADIKVPRLGAKAFQAGFLRI 540

Query: 568 PNAKNFLDNTGVHPESYAVKVLQDITIKELDLAKELQNLNLNATPESIGVGQETLK 627
 P AKN LNTGVHPESY AVK+L L +L+LD AK L + + AE+ +AGETLK
 Sbjct: 541 PNAKNFLDNTGVHPESYAVKVLQDITIKELDLAKELQNLNLNATPESIGVGQETLK 600

Query: 628 DIIEDLLKPGRLDRDQFAPVLRHVDLVDLKVQGLQGTVRNVVDPGA+VDIGVHEDG 687
 DII DLLKPGRLDRDQFAP+LR D+LD+ DL+GQ+L+GTVRNVVDPGA+VDIGVHEDG
 Sbjct: 601 DIIDLLKPGRLDRDQFAPILRDILDLKDLQGLQGTVRNVVDPGA+VDIGVHEDG 660

Query: 688 LIHQSLRIKRRDKKTRKMPLOHPSKLSVGDIVTVVVEVDAERSRIGLSLKP 743
 LIH S + X + HPS+ +SVGD+VTWVW +D +R ++ LSL+ P
 Sbjct: 661 LIHISEMSKTF-----VNHESQVSVGDLVTWVWSKIDLRHKVNLSELP 706

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

50 Example 348

A DNA sequence (GBSx0379) was identified in *S.galactiae* <SEQ ID 1129> which encodes the amino acid sequence <SEQ ID 1130>. This protein is predicted to be N5,N10-methylenetetrahydromethanopterin reductase homolog. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4864 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

-441-

>GP:AAB94650 GB:U96107 N5,N10-methylenetetrahydromethanopterin
 reductase homolog [Staphylococcus carnosus]
 Identities = 164/300 (54%), Positives = 217/300 (71%), Gaps = 1/300 (0%)

Query: 45 VYIGGHHREDFAVSAPRIVLAAGAVRTNNIRLSSAVTILSSNDPIRVYQQFSTIDALS 104
 +YG+GHHR D+AVS P VLAA A T I+LSSAVT+LSS+DP+ VY++F+T+IR+SN
 Sbjct: 1 NYLGGHRSDFYAVSDPVTVLAAAASLTQRIKLSSAVTVLSSDQFVCVYRFPATLAVSN 60

Query: 105 GRAEIMAGRSFIESPFLFGYDLADYDLFNEKMMMLAINSATNLWKGSHITVTVNRFP 164
 GRAEIM GHSFIESPFLFGYDL DYD LF EK++L IN +W+G + +
 Sbjct: 61 GRAEIMVGRSFIESPFLFGYDLADYDLFVEKLKLLKEINQHEVTVWBSITRPAIKGLG 120

Query: 165 IYPRALQRLPIVATQGNVDSTIRIABQGLPIVYATIGGNPKAFQCLVHIYKEVGSRSNG 224
 +YFRA+Q ++PIW+ATGG +S+IR AE GLPI YA IGGNPK F++ + IY+ V G
 Sbjct: 121 YIPRAVQDEIPVILATGGTPESSIRAAEFGLPIYATIGGNPKAFKRNIAITRAVARSNG 180

Query: 225 HKPQLKVAARSHSGWIEEDMGTADIRYFFPTQTVDNIAGRPWSENIKEQYLRSGVGE 284
 + + VA HSG+I + ++ A ++ PTK + IAK R +W T+ + R + E
 Sbjct: 181 YDLADMPVAHSGWGTIADTDEQAGREFYEPTKVHRELIKER-NWPPYTEAHFQREISUE 239

Query: 285 GAIFVGSPEVVAHKIIGLVEALDQFMLHLFVGSMPHKOVAIKLYGEVAPIVKRYF 344
 GA+VFGSPE VA K+I ++E L L+RFVLH+VFGSMH+ ++ AIKLYGK V PI+ YP
 Sbjct: 240 GAMPVGSPEVVARMKIVIEELGLNRFMLHLPVGSMPHERIMKAIKLYGKRVKPIIEDTY 299

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 349

A DNA sequence (GBSx0380) was identified in *S.galactiae* <SEQ ID 1131> which encodes the amino acid sequence <SEQ ID 1132>. Analysis of this protein sequence reveals the following:

Possible site: 22
 >>> Seems to have no N-terminal signal sequence

Final Results

bacterial cytoplasm	---	Certainty=0.1310 (Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>

A related GBS nucleic acid sequence <SEQ ID 9715> which encodes amino acid sequence <SEQ ID 9716> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1133> which encodes the amino acid sequence <SEQ ID 1134>. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence

Final Results

bacterial cytoplasm	---	Certainty=0.0915 (Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 20/40 (50%), Positives = 27/40 (67%), Gaps = 3/40 (7%)

Query: 4 NAITHKNQDLSEMFASPAKVP---KPKKVDSDSKPEQKD 40
 NAITHK+ D+LE N A FA +P KP +V++D K K+

Subjct: 1 MATTHKKNDLEKMLAGFASIPSFDFKLELVNTDGKATKE 40

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

5 Example 350

A DNA sequence (GBSx0381) was identified in *S.agalactiae* <SEQ ID 1135> which encodes the amino acid sequence <SEQ ID 1136>. Analysis of this protein sequence reveals the following:

```

Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1453(Affirmative) < succ>
bacterial membrans --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

20 Example 351

A DNA sequence (GBSx0382) was identified in *S.agalactiae* <SEQ ID 1137> which encodes the amino acid sequence <SEQ ID 1138>. Analysis of this protein sequence reveals the following:

```

Possible site: 37
>>> Seems to have an uncleavable N-term signal seq

INTEGRAL    Likelihood = -11.15    Transmembrane 216 - 232 ( 210 - 240)
INTEGRAL    Likelihood = -9.18    Transmembrane 15 - 31 ( 10 - 39)
INTEGRAL    Likelihood = -9.02    Transmembrane 283 - 299 ( 276 - 299)
INTEGRAL    Likelihood = -8.76    Transmembrane 128 - 144 ( 119 - 150)
INTEGRAL    Likelihood = -4.62    Transmembrane 243 - 259 ( 237 - 265)
INTEGRAL    Likelihood = -2.44    Transmembrane 65 - 81 ( 65 - 81)
INTEGRAL    Likelihood = -2.44    Transmembrane 94 - 110 ( 93 - 111)

----- Final Results -----
bacterial membrane --- Certainty=0.5458(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:CA812119 GB:Z99105 ycgR [Bacillus subtilis]
Identities = 141/283 (49%), Positives = 198/283 (69%), Gaps = 3/283 (1%)

Query: 10 SVLQNFAPFISIIIALPFVLLGTLISGIIIVFITPDIVNKKFLPKNFILVLEFGTFGVGV 69
      S L Q + I P I S I + I R A + P F + L + G I L S G I I + F + + + + + + P K N + F L V L F G G +
      S F L Q L N S I P I S I L I E A I P F I L G I V L I S G I I Q M F V S E E M I A R I M P K N R F L A V L F G A L A G V L 65
Subjct: 6

Query: 70 F P S C E G I I P I I N R F L E K K V P S Y T A V P F L A T A P I I N P I V L F A T Y S A F G N S I R F I L R F P V G 129
      F P + C E G I I P I R L K V P + V F + T A P I I N P I V L F + T Y A F G N + R
      F P A C E G I I P I T R L L K G V P L H A G V A M L I A P I I N P I V L F S T Y I A P N R G S V V F Y R G L 125
Subjct: 66

Query: 130 A T T V A I A L G V M L A P L V D N I L K E D A K P T H P M D Y S D K K W Y Q K I F A L A H A I D E F F D T G R Y L 189
      A V + + G V + L + + D N L + + P H R + + Q K + L H A I D E F F G + Y L
      A L A V S L I I G V I L S Y G F Q D N Q L L K P D E P G H I I H H K T L - L Q L G G T L R H A I D E F F S V G Y L 184
Subjct: 126

Query: 190 V P G T L I A S A M Q I V L P T R V L T T I G S P I T A I L V M M L A P I L S L C S E A D A F I G A S L L S T P G I 249

```

-443-

+ G IA+AMQ Y+ T L IG + +++ LVM LAF+LSLCS DAFI +S STF +
 Sbjct: 185 IIGAFIAAMQYTVKTSLLAIGQNDVSSSLVMGLAFVLSLCSVDAFIASSFSSTPSL 244

Query: 250 AFVLAFLIGMVDIKNLMMVAKPKRFIVQFVSVSLII 292
 ++AFL+ G MVDIKNL+MM+ AFK RF+ F+ ++ ++I+
 Sbjct: 245 GSLIAFLVFGMVDIKNLMMLAFAKGRFV - FLITTYIVIVY 285

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1139> which encodes the amino acid sequence <SEQ ID 1140>. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -9.92	Transmembrane	216 - 232 (211 - 237)
INTEGRAL	Likelihood = -9.45	Transmembrane	283 - 299 (276 - 299)
INTEGRAL	Likelihood = -8.76	Transmembrane	128 - 144 (119 - 150)
INTEGRAL	Likelihood = -7.80	Transmembrane	15 - 31 (10 - 39)
INTEGRAL	Likelihood = -5.47	Transmembrane	243 - 259 (237 - 265)
INTEGRAL	Likelihood = -2.44	Transmembrane	65 - 81 (65 - 81)
INTEGRAL	Likelihood = -2.44	Transmembrane	94 - 110 (93 - 111)

20 ----- Final Results -----
 bacterial membrane --- Certainty=0.4970 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

25 The protein has homology with the following sequences in the databases:

>GP:CB12119 GB:Z99105 ycgR [Bacillus subtilis]
 Identities = 143/288 (49%), Positives = 196/288 (67%), Gaps = 1/288 (0%)

Query: 10 SVLQMFALPMSIIIEALPFVLLGTLGSCIEVFTPLVQKYLKQKICRLRGTFVGFV 69
 S LQ +IF+SI+IEA+PF+L+G ILSG I+V+V+ E+ + +PK + L +LFG G +
 Sbjct: 6 SFLQNSIPISIIIEALPFIILGVLGSGIIQMFVSEMIARIMPKNRFLAVLFGALAGVL 65

Query: 70 PFSCCGIIPINRPLEKKVPSYTAVPFLATAPINPIVLPATYSAPGNSLRLIRLVG 129
 PF+CECGIIP I R L K VP + V F+ TAPINPIVLE+TY APGN + R
 Sbjct: 66 PFACCGIIPITRRLKLVQVLEAGVAFMLTAPINPIVLPATYSIAPGNRNSVVFVGGGL 125

Query: 130 AALVAITLGVMLAFIVDDNLIKNAQVFHFHDSHESLPKRIYALVHAIDFPTGGRYL 189
 A V++ +GV+L++ DN L +P H H + H L +++ L HAIDFPP G+YL
 Sbjct: 126 ALAVSLIIGVLSYQFQDNQQLKPDPEGH-HHHHGTLLQKLGTLHAIDFSPVSGYL 184

Query: 190 VEGTLIASAMQIVTVKTSLLAIGQNDVSSSLVMGLAFVLSLCSVDAFIASSFSSTPSL 249
 + G IA+AMQ YV T L IG N +++ L+MM +AF+LSLCS DAFI +S STF +
 Sbjct: 185 IIGAFIAAMQYTVKTSLLAIGQNDVSSSLVMGLAFVLSLCSVDAFIASSFSSTPSL 244

Query: 250 AFVLAFLIGMVDIKNLMMVAKPKRFIVQFVSVSLIAVCLIV 297
 ++AFL+ G MVDIKNL+MM+ AFK RF+ I V+++ LLV
 Sbjct: 245 GSLIAFLVFGMVDIKNLMMLAFAKGRFVFLITTYIVIVLAGSLIV 292

An alignment of the GAS and GBS proteins is shown below:

Identities = 248/300 (82%), Positives = 278/300 (92%)

Query: 1 MDIFNQLPDSVLQWFAIFISIIIEALPFVLLGTLGSCIEVFTPLVQKYLKQKICRLRG 60
 M +F+ LP SVLQMFALP+SIIEALPFVLLGTLG IEVF+TP++V K+LEK K LR+
 Sbjct: 1 MSFLPSNLVPSVLQWFAIFMSIIIEALPFVLLGTLGSCIEVFTPLVQKYLKQKICRLR 60

Query: 61 LGTFVGFVFPSCCGIIPINRPLEKKVPSYTAVPFLATAPINPIVLPATYSAPGNSL 120
 LGTFVGFVFPSCCGIIPINRPLEKKVPSYTAVPFLATAPINPIVLPATYSAPGNS+
 Sbjct: 61 LGTFVGFVFPSCCGIIPINRPLEKKVPSYTAVPFLATAPINPIVLPATYSAPGNSL 120

Query: 121 RFLRLRWGATVIAIALGVMLAFIVDDNLIKNAQVFHFHDSHESLPKRIYALVHAID 180
 RFLRLR WGA +VAI LGVMLAF+VDDNLIK+ +A+P HFHDS + + +I+LAL HAID
 Sbjct: 121 RFLRLRWGAALVAITLGVMLAFIVDDNLIKNAQVFHFHDSHESLPKRIYALVHAID 180

Query: 181 EFPDTRGLYVFGTLIASAMQIVLPTVLTITGHSPITALVMMGLAFVLSLCSVDAFI 240

-444-

```

      EFPDTRGYLVFGTLIASAQIY+PTRVLTTIGH+P+TAIL+MMI+AFILSLCSEADAFIG
  Subjct: 161 EFPDTRGYLVFGTLIASAQIYVPTRVLTIGHNPLTAILMMMAFILSLCSEADAFIG 240
  Query: 241 ASLLSTFGIAPVMAFLILGPMVDIKNLA*MMVNSFKTRFIVQFISVSSLIITIIYCLFVGVI 300
          ASLLSTFG+APV+AFILIGPM+DIKNL*MMV +PK RPIVQFI VS L+I +YCL VGV+
  Subjct: 241 ASLLSTFGVAPVLAFLILGPMVDIKNLA*MMVKAFGKRPVQFIGVSVLMAVYCLLAVGL 300

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 352

A DNA sequence (GBSx0383) was identified in *S.agalactiae* <SEQ ID 1141> which encodes the amino acid sequence <SEQ ID 1142>. Analysis of this protein sequence reveals the following:

```

  Possible site: 13
  >>> Seems to have no N-terminal signal sequence

  ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.4703(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25 Example 353

A DNA sequence (GBSx0384) was identified in *S.agalactiae* <SEQ ID 1143> which encodes the amino acid sequence <SEQ ID 1144>. Analysis of this protein sequence reveals the following:

```

  Possible site: 50
  >>> Seems to have an uncleavable N-term signal seq

  INTEGRAL    Likelihood = -8.44    Transmembrane    45 - 61 ( 39 - 65)
  INTEGRAL    Likelihood = -8.12    Transmembrane    83 - 99 ( 77 - 101)
  INTEGRAL    Likelihood = -0.00    Transmembrane    2 - 18 ( 1 - 19)

  ----- Final Results -----
          bacterial membrane --- Certainty=0.4376(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8559> which encodes amino acid sequence <SEQ ID 8560> was also identified. Analysis of this protein sequence reveals the following:

```

  Lipop Possible site: -1    Crend: 2
  SRCFLG: 0
  MCG: Length of UR: 8
          Peak Value of UR: 2.23
          Net Charge of CR: 1
  MCG: Discrim Score: 0.46
  GVH: Signal Score (-7.5): -3.54
          Possible site: 42
  >>> Seems to have an uncleavable N-term signal seq
  Amino Acid Composition: calculated from 1
  ALLOW program    count: 2 value: -8.44 threshold: 0.0
          INTEGRAL    Likelihood = -8.44    Transmembrane    37 - 53 ( 31 - 57)
          INTEGRAL    Likelihood = -8.12    Transmembrane    75 - 91 ( 69 - 93)
          PERIPHERAL  Likelihood = 2.76    200

```

modified ALON score: 2.19
icml HYPID: 7 CFP: 0.438

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.4376 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CA12118 GB:Z99105 yegQ [Bacillus subtilis]
Identities = 100/290 (34%), Positives = 159/290 (54%), Gaps = 25/290 (8%)

Query: 9 MIRFLILAGYFELSMYLLKSLKINQYINHTYTLAYISMVLSFILAIVQLIMVNMGMH 68
N R L+L G+ +L SG L +YIN Y YL++I++ L IL VQ ++K+ +
Sbjct: 1 MRFLLVLMGFTFFFYHLHAGSNLTKYINMKYALSFIALFLLAILTAVQAYLPKSPK 60

Query: 69 SHLRGKIA-----KSTSP-----NILVFPVLVGLLVPTVSLDSTTSVAKGYN 110
H H + P ++ +PP++ G+ P +LDS+ V KG++
Sbjct: 61 GHRHHDHCOGOGHDEHDEHQBKPFYQRYLYIVVFLPVLVSGIFPPFIATLDSIVKTKGFS 120

Query: 111 FPLAAGSTGTVSGDQTRVQYLKPTSTYFTSSAYEKMQRLEKKGKSGTLITITENYME 170
F A S SQ QYL+PD S Y+ +Y+K+M++ KY ++T +++++
Sbjct: 121 FK-AMSGSDHYSQ----TQYLRPDASLAYACDSYDKMKQLFNKYSSKKEISLTDODFLK 175

Query: 171 VMELIYLYSQFMDRQIQITGPFYI-NEPKHGVQPIFRPGIITHCIADSGVYGLLT-GNQ 228
ME IY YP +F+ R I++ GF Y ++ P+ RFGIITHCIADSGVYGL+L
Sbjct: 176 GMETITNYVGEPLGRITIEFGFAYKGNAINKQLVLRPGIITHCIADSGVYGLMVEFPKD 235

Query: 229 KSYFNTNVTVVGRTIKSEYNQLLQNLFLVHIESRQVSKANNFYVYVF 278
D+ N+ +GT+ SEY Q + LFV+ + + + K ++FYVIR F
Sbjct: 236 MDIKDDENIHINGTLASEYYQPFKSTLVVVKVTDWNIKKPDDPYVYKGF 285

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1145> which encodes the amino acid sequence <SEQ ID 1146>. Analysis of this protein sequence reveals the following:

Possible site: 60

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -8.33 Transmembrane 83 - 99 (74 - 101)
INTEGRAL Likelihood = -6.21 Transmembrane 42 - 58 (39 - 62)

----- Final Results -----

bacterial membrane --- Certainty=0.4333 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9115> which encodes the amino acid sequence <SEQ ID 9116>. Analysis of this protein sequence reveals the following:

Possible cleavage site: 54

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -8.33 Transmembrane 75 - 91 (66 - 93)
INTEGRAL Likelihood = -6.21 Transmembrane 34 - 50 (31 - 54)
PERIPHERAL Likelihood = 2.76

----- Final Results -----

bacterial membrane --- Certainty= 0.433 (Affirmative) < succ>
bacterial outside --- Certainty= 0.000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty= 0.000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 208/279 (74%), Positives = 244/279 (86%), Gaps = 1/279 (0%)

Query: 1 MFIOGGMIRFLILAGYFKLSMYLKLSGKLNQYINHYTYLAYISMVLSFLIALVQLII 60
 +F CGG +MIRFLILAGYFEL+MYL+LSGKLG+QYIN Y+YLAYISM+LSFLIA+VQL
 Sbjct: 1 LPTGGALMIRFLILAGYFEL+MYL+LSGKLG+QYIN+VRYSTYLAYISMVLSFLIALVQLIT 60

5 Query: 61 WVKMKQSHSLHGKIAKSTSPMLVFPVVLVGLLVPTVSLDSTTVSANGYFPLAAGSTGT 120
 W+K+K+HSHL GKIA+ TSP ILVFPVL+GLVPTV+LDSTTVSANGY FPLAAG++ T
 Sbjct: 61 WMONIKVSHSLHGKIALRTSPFVLVFPVVLVGLLVPTV+LDSTTVSANGYFPLAAGAKST 120

10 Query: 121 -VSQDGTIRVQYLKPDSTSYFTISSAYEKRMQKELKKYKSGTILITITENYMEVMEILIYLP 179
 VS DGT +QYLKPDTS YFT SAY+KEM++EL KYNG +TITENYMEVMEILIYLP
 Sbjct: 121 GVSQDGTITQYLKPDSTSYFTKSAYQKEMRQELKRYKGGKPVITITENYMEVMEILIYLP 180

15 Query: 180 EQFMDRQIQYTGFPVYNEPKHEGYQPIFRFGIIHCIAADSGVYGLLTQNGQSYPDNTWTV 239
 ++F+DR IQYTGFPVYNEP H+ YQF+FRFGIIHCIAADSGVYGLLTQNGQ SYP+NTW+TV
 Sbjct: 181 DEFLDRDIQYTGFPVYNEPCHDNYQLFRFGIIHCIAADSGVYGLLTQNGQSYPDNTWTV 240

Query: 240 RGTINSEYNQLLQNLVFLMHIESRQVSQANFYVRVP 278
 +G + EY++ L+Q+LFLV + E Q + NNFYVRVP
 Sbjct: 241 KGRLEMYEDKNLEQLFLVQLAELVQITKEPNNFYVRVP 279

20

SEQ ID 8560 (GBS235d) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 146 (lane 14 & 15; MW 48.5kDa). It was also expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 146 (lane 17 & 18; MW 23.4kDa), in Figure 150 (lane 15; MW 23kDa) and in Figure 182 (lane 5; MW 23kDa).

25 GBS235d-His was purified as shown in Figure 235, lane 6-7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 354

A DNA sequence (GBSx0385) was identified in *S. agalactiae* <SEQ ID 1147> which encodes the amino acid sequence <SEQ ID 1148>. This protein is predicted to be signal recognition particle (ftsY). Analysis of this protein sequence reveals the following:

Possible site: 57
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3301 (Affirmative) < succ>
 bacterial membrans --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

40 The protein has homology with the following sequences in the GENPEPT database:

>GP:BA06205 GB:AP001515 signal recognition particle (docking
 protein) [Bacillus halodurans]
 Identities = 175/304 (57%), Positives = 227/304 (74%)

45 Query: 233 EKYNSLAKTCTGFSARLNAFLNFRVRDERTFPELEEMLLSDGVGVNATVQLTDLAYE 292
 EK+ L+KTR F+ ++N + +R VDE+FFELEEL+LI +DWGV L B+L+ E
 Sbjct: 20 EKFAGLEKTRDTSFAGKMDLVYKRSVDEDFPELEESTLIGADVGVTVMVMDLVELKDE 79

50 Query: 293 AKLENAKKSDELKRVIVEKLVEIYEKDIYNEADINPQGLAVMLFVGKNGVKTSIGKL 352
 + +N K S D++ +I EKL E+ EK+ G E GL+V+L VGVKNGVKTSIGKL
 Sbjct: 80 VRQKHLSKDIQPIISEKLAELLEKGGSETVNLQPSGLEVLVGVGVKTSIGKL 139

Query: 353 AHQYKQSKKVMVAADTFRAGAVAGLVEKGRVVDVFPVVTGEEKADPASFVDEGKQVA 412
 AH YK GSKK+L A ITFRAGA+ QL WG R V V+ E +DR+V+FD ++A +
 Sbjct: 140 AHYKQKQSKKVLAMGDTFRAGAIQLEWGERAGVDVIKQSESGDPANVMDTAQAKS 199

55 Query: 413 GQVDVLLIDTAGRLQNKENIMAELEKIGRIKRVVDPAPHEITLALDASTGQNALSQAK 472

-447-

+ D+L+ ITAGRLQNK NIM ELKQ+ R+I R +P APHE L+ALMA+TGQNA+SQAK
 Sbjct: 200 READLILCITAGRLQNKVIMMKLEKVEKVRVISEITPGAPHEVLIALDATTGQNAHSQAKT 259
 Query: 473 PSKITPLTGLILTKIDGTAKGGVVLAIROELDIPVKFIGFGKEIDDIGPNSDEFMRGL 532
 5 F + T +TG+ILTK+DGTAKGG+VLAI R ELDPVKF+G GEKIDD+ P+SE P+ GL
 Sbjct: 260 PKETITDVTGLILTKIDGTAKGGVVLAI RHELDIPVKFVGLGEKIDDLQPPDSBQFVGLF 319
 Query: 533 EGIL 536
 10 + ++
 Sbjct: 320 KDMV 323

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1149> which encodes the amino acid sequence <SEQ ID 1150>. Analysis of this protein sequence reveals the following:

Possible site: 60
 15 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4384(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 20 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 339/549 (61%), Positives = 404/549 (72%), Gaps = 46/549 (8%)

25 Query: 1 NGLFDRLPGHKKKKEPEIEASESVLEDESDVIDKEGNSFKESTLNRTSSEVPVAEDD 60
 NGLFDRLPG K+ K E+ E+++ E KER S + E ++ + +
 Sbjct: 1 NGLFDRLPGKIKETPKVLESEKLEENLITE----TTQKEELSEKANEQ-----DKIEAVQEQ 51
 30 Query: 61 SFLELSRDTALEBSHQVPTSEIHPLSESDTSEIPVKEDDSFLELEDRAKTKVADTSEVIN 120
 ++ + A S + P + ++ L E+T D + IT+E
 Sbjct: 52 ---DVSSEGCAGSEVNGFEAASVNALVSEETG-----DENSEHPSSEITNEF-- 92
 Query: 121 VVPDSITLSDNVAKSEASPSDKBQLEDSQAQSDQPSSETPLQSEMS--SGKTEVQTSSEDT 178
 D T L VS S++ S+ + L D +QP Q + S S E S++
 35 Sbjct: 93 -AADKTDLK--VEELQSTASEPKDLVDQPVVEQFPPTKQACADASHDSANEAVDTSEKQ 149
 Query: 179 SAADAPLADYYAKRKAIKEHLSNELST-----DESEFSEAQVLEQSQCA--DTIK 227
 S++ + DTY ++ A+EK + + +T R++ S + E SQ++A DTI
 Sbjct: 150 SSEQQVMEDYYRRGAALKSLQKAAATVPVMPSEVPQENQASTSAA--SQNKATHTTIP 208
 40 Query: 228 AESQEEKYNRSILKKTITGPSARLNAPLNPRVDEEFPSELEMLILSDGVGNVATQLTE 287
 E+ +EKY RSLIKKTITGPSARLN+P +NPRVDKRPFR+LEEMLILSDGVG+VAT LTE
 Sbjct: 209 -ETDQEKYKRSILKKTITGPSARLNPFANPRVDEEFPDELEMLILSDGVGNVATILTE 267
 45 Query: 288 DLRYEAKLENAKKSDELKRVIVEKLVIEYKDGIVNEALNFOEGLTVMPLVGVNGVCKIT 347
 +LRYEAKLENAKK + LKRVIVEKLV+IYKDG VNEALN+Q+GLTVMPLVGVNGVCKIT
 Sbjct: 268 ELRYEAKLENAKKDALKRVIVEKLVVDIYKDGIVNEALNHYQGLTVMPLVGVNGVCKIT 327
 Query: 348 SIGKLAHQYKSGKVMVLAADTFRAGAVQLVNGRRVDVDPVVTGSEKADPASVVDGM 407
 SIGKLA++YK +GKVMVLAADTFRAGAVQLVNGRRVDVDPV+TG EKADPASVVDGM
 50 Sbjct: 328 SIGKIAYRYKQSGKVMVLAADTFRAGAVQLVNGRRVDVDPVITGPEKADPASVVDGM 387
 Query: 408 EKAVAGQVDVLLIDTACRLQNKENIMAEKELKIGRIIKRVDPDAPHETLLALDASTGQNAL 467
 EKAVA+GVD+LLIDTACRLQNKENIMAELEK+GRIIKRV+PDAPHETLLALDASTGQNAL
 55 Sbjct: 388 EKAVAGQVDVLLIDTACRLQNKENIMAEKELKIGRIIKRVDPDAPHETLLALDASTGQNAL 447
 Query: 468 SQAKEFSKITPLTGLILTKIDGTAKGGVVLAIROELDIPVKFIGFGKEIDDIGPNSDEF 527
 SQAKEFSKITPLTGLILTKIDGTAKGGVVLAIROELDIPVKFIGFGKE+DDIGRP+SEDF
 Sbjct: 448 SQAKEFSKITPLTGLILTKIDGTAKGGVVLAIROELDIPVKFIGFGKVIDDIGPNSDEF 507
 60 Query: 528 MRGLLEGIL 536
 M+GLLEGIL
 Sbjct: 508 MKGILLEGIL 516

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 355

A DNA sequence (GBSx0386) was identified in *S. galactiae* <SEQ ID 1151> which encodes the amino acid sequence <SEQ ID 1152>. Analysis of this protein sequence reveals the following:

Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3592 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA62048 GB:L10328 f270 [Escherichia coli]
Identities = 101/273 (36%), Positives = 160/273 (57%), Gaps = 10/273 (3%)

Query: 4 IKILALDLDGTLPTTDKKVSEENKVALKAAREKGIKVVITGRPLKAIGNLLEDLVSD 63
IK++A+D+DGLT D +S K A+ AAR +G+ VV+TTGRP + N L++L +
Sbjct: 3 IKLIAIMDGTLLPDHTISPAVISAIAAARGSVNVITGRPYAGVHNYKLKLMSPQ 62

Query: 64 EDSYITFNGGLVQONT-GKILAKTAMTROVEDIHEELVQGLPTDILSGTVIS----I 118
DY IT+NG LVQ+ G +A+TA++ + + + +VG L T Y+ I
Sbjct: 63 GDYCITYNGALVQKAADGSTVQAQTAISYDOYRLEKLSREVGSHFALDRTLLTANRDI 122

Query: 119 ANKGHSHOYHLANPLLEFIEVDLEQVFKDVVYNKIVSVIDATYLDQIAKLPRLKVDY 178
+ H + PL+ F E B++ + + K++ + + LDQ IA++P +K Y
Sbjct: 123 SYTYTHESPVATIPLV-FCEA---HMDPNTQFLKVMIDEPAILDQALARPQXVKEKY 178

Query: 179 ENFKSRDILILEMPKGVHKAAGLELLTGHGLDSSQVMAMQDEANDLSMLEAGLVANA 238
+ KS LS++ K V+K G++ L LG+ ++MA+GD+ ND++M+E+AG+GVAM
Sbjct: 179 TVLKSAPYFLLELDKRVNKGTVKSLADVLGIKPBREINAGDQENDIAMETAGVGVAND 238

Query: 239 NGIPEAKAIKATTCNNDESQVVAEAIKYLIS 271

N IP K+A T +N E GVA AI KY+L+
Sbjct: 239 NAIPSVKEVANFVT-KSNLEDGVAFKIKVLN 270

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1153> which encodes the amino acid sequence <SEQ ID 1154>. Analysis of this protein sequence reveals the following:

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3502 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 180/273 (65%), Positives = 218/273 (78%), Gaps = 1/273 (0%)

Query: 3 DIKILALDLDGTLPTTDKKVSEENKVALKAAREKGIKVVITGRPLKAIGNLLEDLVSD 62
+I+ILALDLDGTL+ T+K V++ NK AL AAREKG+KVVITGRPLKAIGNLLE+I++L+
Sbjct: 2 NIRILALDLDGTLTYNTEKIVIDANKKALAAREKGVKVVITGRPLKAIGNLLEDLVSD 61

Query: 63 DEDYSITFNGGLVQONTGKILAKTAMTROVEDIHEELVQGLPTDILSGTVISANK- 121
+DYSITFNGGLVQNTG+NYG++L K++++ +V I + L VGLPTDI+S G VYSI +K
Sbjct: 62 HDDYSITFNGGLVQNTGKIVLKDSSLSFDVQVQIQQAELVGLPTDIISGDVYSIPSPK 121

Query: 122 GHHSQYHLANPLLEFIEVDLEQVFKDVVYNKIVSVIDATYLDQIAKLPRLKVDYEMF 181

-449-

G HSQYHLANPL FIEV ++PKD+ YNKIV+V D +LQOI KL L D+E F
 Sbjct: 122 GRHSQYHLANPLITFIVTSVABLEPKDITYNKIVTVTDPDFLDQOI KLSPSLDFEDFRAF 181

Query: 182 KSRDIIILEMPKGVHQVGLLELTUHLGLDSSQVMAMGDEANDLSMLHAGLVAMANGI 241
 KSRDII E+PKG+ KA GL LL +HLGLD+ VMAMGDEAND +MLEWAGLVAMANG+
 Sbjct: 182 KSRDIIIFIMPKGIDKAFGLMLCQHLGLDRAHVAMGDEANDFMLEWAGLVAMANGV 241

Query: 242 PEAKAIKATTICNNDSSGVAEAGKYLSEEN 274
 AKA A A T NDSGVAEA+ +IL EE+
 Sbjct: 242 SGAKADADAVTTLNDSGVAEAVKTFILSEES 274

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 356

- 15 A DNA sequence (GBSx0387) was identified in *S.galactiae* <SEQ ID 1155> which encodes the amino acid sequence <SEQ ID 1156>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4648(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the GENPEPT database:
 >GP:BA355556 GB:D90723 Hypothetical 30.2 kd protein in idh-deoR
 intergenic region. [Escherichia coli]
 Identities = 91/264 (34%), Positives = 146/264 (54%), Gaps = 4/264 (1%)

30 Query: 2 IKLVATDMDGTFLDENGTYDKKRLANVLKFKKEQGVFTAAAGSRSLSELEQLADFRDM 61
 IKLA A DMDGTFL + TY++R ++ K QGI F ASG L F + +++
 Sbjct: 4 IKLIVATDMDGTFLSDQKTYNRERPMQYQGMKQAGIRFVVASGNQTYQLISFFFEIANEI 63

35 Query: 62 AFVARENGSAAVLFHRLAYEQHLSRECYLDIIDLKSKSPYMNWETVLSKGDGAYLLGDAN 121
 AFVARENG V + + LS++ + ++HL F + E + GK+ AY L +
 Sbjct: 64 AFVARENGHVVSEKDVFNGLSKDAFATVVEHLLTRFEV---ETIACGKNSAYTLKKID 120

Query: 122 FDIYIRFITHYDNLQKVSHPEDVDIIKFTVNFANFETSTVROAESWNAI-PIYATVTTG 180
 YY L+ V +P+++DI FK N ++E + Q ++ ++H+ +V TG
 Sbjct: 121 DAMKTVAEYTHRLKYVDNFDNLEIDFFKFGMLNDELIPQKALHEAIGDIMVSVHTG 180

40 Query: 181 FKSIDITLSSVNRKNGLEHLCEQYGIRAEVLSFGDNINDLEMLWGGKAIATENARPEV 240
 SID+I+ V+K NGL L + +GI EV+ PGD ND+EML +Q + A ENA V
 Sbjct: 181 NGSIDLIIPGVHKNKRLQ/LQKLGIDDSVVVFGDGGNDIEMLRQAGSFVAMEKASAV 240

45 Query: 241 KEIADCTIGHNNQAVAYLESNV 264
 A G +N + V+ ++ ++
 Sbjct: 241 VAAAKYRAGSHNRKGVLDVIDKVL 264

- 50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1157> which encodes the amino acid sequence <SEQ ID 1158>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3401(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 60 An alignment of the GAS and GBS proteins is shown below:

-450-

Identities = 138/265 (52%), Positives = 193/265 (72%), Gaps = 1/265 (0%)

Query: 1 MIKLVAATMDGTFLDENSTYDKKILANVLKKFKKEGIVPTAASGRSLSLLEQLPADFRDQ 60
 5 Sbjct: 1 MIKLVAATMDGTFL E+GTY++++LA +L K E+GI+F +SGRSL++++QLF F DQ 60
 MIKLVAATMDGTFLAEDGTYNQSLAALLPKLAEGILFAVSSGRSLLAIDQLFEPFLDQ 60

Query: 61 MAFIAENGSAAVLNLRLAYEQHLSREQYLDI IDHLSKSPYMENNEYVLSGDKGAYILSDA 120
 10 Sbjct: 61 IAVIAENGSSVVQYRGRILFADHMTKEQYTEVAKKILANEHYVETGMVPSGQKAAYILKGA 120
 +A IAENGSS + + ++EQY + + + +P+ V SG+ AYIL A

Query: 121 NPOYIEPTHTYIDNLQKVSHFEDVD-DLIFKVTANFTFETVRQAEZBWNAIPYATAVTT 179
 + +YI+ HYY N++ ++ FRED+ D IFKV+ NFI TV + +W+NQA+PYATAVTT
 Sbjct: 121 SEBYIQKTHYYANVKVINGFEDNENDALFKVETNFAGHTVLEGSQWLNALPYATAVTT 180

Query: 180 GFKSIDILLSSVNRKNGLEHLCSGYIRAEVLSFGDNDNLEMLEWSGKALATENARPE 239
 15 Sbjct: 181 GF SIDIIL VNK G+EHLC+ GI+ E ++FSDN ND +MLE++G+ALATENARPE 240
 GFSDIDIILKEVNRNGFMEHLQALGIIKGAETIAPGDNFNDIQMLEFAGRAATENARPE 240

Query: 240 VKETADCTIGHHNAQWAVYLSNV 264
 20 Sbjct: 241 IKVISDQVIGHCNDGAVLTYLKLGV 265

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 357

A DNA sequence (GBSx0388) was identified in *S. agalactiae* <SEQ ID 1159> which encodes the amino acid sequence <SEQ ID 1160>. Analysis of this protein sequence reveals the following:

Possible site: 18
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2428 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

35

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 Example 358

A DNA sequence (GBSx0389) was identified in *S. agalactiae* <SEQ ID 1161> which encodes the amino acid sequence <SEQ ID 1162>. This protein is predicted to be p115 protein (smc). Analysis of this protein sequence reveals the following:

Possible site: 55
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -4.99 Transmembrane 1092 -1108 (1088 -1110)

45 ----- Final Results -----
 bacterial membrane --- Certainty=0.2996 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

50

A related GBS nucleic acid sequence <SEQ ID 9713> which encodes amino acid sequence <SEQ ID 9714> was also identified.

-451-

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB13467 GB:Z99112 chromosome segregation SMC protein homolog
[Bacillus subtilis]
Identities = 458/1193 (38%), Positives = 728/1193 (60%), Gaps = 27/1193 (2%)

5 Query: 1 MFLKEIEVQGFKSPADTKTVFDDGVTVVGFVGSCKSNITSLRWALGSSSAKSLRGK 60
MFLK +++ GFKSTAA++ V+V +GVTVVGFVGSCKSNIT++RW LGE SA+SRLGSK
Sbjct: 1 MFLKRLDVGFKSFAERISVDVFKGVTVVGFVGSCKSNITDAIKWVLGQSARSRLGSK 60

10 Query: 61 MEDVIEAGTBNKPNLNAQVSTVLNDSHFPIKNIADSVRVERRIFFNKDLSYLLIDGRKVR 120
M D+IFAG++R+LN A+V++TLDN DHE+ EV V RR++R+G+SB+LI+ + R
Sbjct: 61 MEDITFAGSDSRKRLNIAEVTILTLDDDDHFLPIDFHEVSVTRVRKSGSEFLINQPCR 120

15 Query: 121 LRDIIDLDMUTGLGRDUSFIIISQGRVEAIFPSKPEERRAIFEEAGVLAHYTKTKKETQSK 180
L+DI DLPM+DLG+++FSIIISQ+VE I +SK E+RR+IFEEAGVLAHYTKIRKK+ ++K
Sbjct: 121 LKDIIDLDMDSGLGKAFSIIISQKVEIISLSSKEDRRSIFEEAGVLAHYTKRKKAENK 180

20 Query: 181 LEQTQGNLDRLEDIIYELDMQVQPLEKQASIAKRFNLDEERQGLHLSILLEDLHQSD 240
L +TO NL+R+EDIA++EL+ Q+PL+ QASTAK +L +E + + + + + DI +
Sbjct: 181 LPEQTQGNLNRVEDILHELEQVLEPLKIQASIAKDYLEKKKLEHVEIALTAYDIKKLHAK 240

25 Query: 241 LITVEKLLTVRKLELITYYQKQSLDEQNSLKQKRHHLSSEEIAKGLALDVTKLKSDL 300
+T++EK+ ++E +E++ +K LE+ + Q LL+ L L
Sbjct: 241 WSTLKEKVVQAKRELAESSAISAKAKIEI/TRDKIALDESVMELQVLLVTSSELEKL 300

30 Query: 301 ERQIDLIRLESNQAEKKEEAGQRLALEIKAKDCSDQITQKNIELTTLSEKIAQIRSEI 360
E + + + + + ++E+ + + + + K + + + + + TL + Q + + + + +
Sbjct: 301 EGRKEVLKERRKQNAVQMQBLEAIVQFQKKEVLEKESLQEAVFETLQAEVQKRLQV 360

35 Query: 361 VSTSSLEFSTFNDPQIIIEKLEDFVTLMQEEADTSNALTALLADINQKQASQAKSRI 420
+ +L + N + + IE+L+ D+ L+ +A N L L+ D+ Q + + + +
Sbjct: 361 KEKQALSLNHEHVEEKIEQLKSUYPELLNSQASIRNEL-QLLDDQMSQSAVTQLRLDN 419

40 Query: 421 QEVSKILEVLSHNKVALE-RFEAAKNVQRLLSHYDQLOQTQLNLQLEHFNQKQSILFDH 479
E S K A E F ++ + + Y+D+ + + + + +S L+
Sbjct: 420 NEKHLEQRHDISARKACETEFARIBQRIHSQVQAYRDMQTKYEQKQRYEKNSLSALYQA 479

45 Query: 480 LDEIKSKQARISLESILKNSHFYAGVSVLQAKDLQGIIGAVSEHLSFQKHYYQTALR 539
++ + + + LE++ + S FY GVK VL+AK++LGGI GAV E++ ++ Y+T+E
Sbjct: 480 YQTVQAKSKKMLSTMQDPSGPGYQGVKEVLKAKERLGGIRGAVLELSTBEQKYETAIE 539

50 Query: 540 IALGSSSQHIIVEDESAARSLAPLKKINROGRATFLPLTTIKPRELAQHYLSKLQSSQG 599
IALG S+QH++ +DE +A++I+L+K+N GRATFLPL+ I+ R+L F
Sbjct: 540 IALGASQHVVTDDQSARKAIQYLKQNSFGRAFLPLSVIRDRQLGSDAETAAHRSSP 599

55 Query: 600 LGIASLVTYDQRLINIPKNNLGLTALFDTVDNANVAARQLNQVRLVLTDDOTELRPOGS 659
LG+ASEAT+D ++ +N LG I + + AN A+ L + + R+VTL+G + PGGS
Sbjct: 600 LGVASELVTDFPAYRSVIQNLTLTTLITEDLGANELAKLLGHRYRIVTLBQDVVNPOGS 659

60 Query: 660 YSGGANRQNTVFI--KPELDNLKKEKLAQCAQSKQLIQHEVATLLEQLKEQETLQALKN 717
+GA ++ N + EL++ K L + + K + E+EV TL ++ ++ ++ I+A L+
Sbjct: 660 MTGGAVKKNNLSLGRSRELEVDVTKRLAERREKTALELQGVTLKHSIQDMEKKLADLRS 719

65 Query: 718 DQEQARLEBORADIEYQQLSEKLADLNKLYGLQSSGALECTTSENE--KNRLEKLEQ 775
GE RL++Q + +L ++N AL ++ E + K +L+EL
Sbjct: 720 TGEGLKLQGVQKGLYELQVAKRNINTHLELYDQEKSALESDESDEKVKRKRKLRELSA 779

Query: 776 FAIKKEELTSLAQIKEDKSIQKVNNNITLSESAQLEERDLIANEQKFRANCTR--- 832
+ K +L I I ++ + K + +L+ L+E ++ K E N RL
Sbjct: 780 VSEKMKOLESDIDRLTKQKQTSSTKESLSNELTELKIAAKKEACGSEEDNLARLKE 839

Query: 833 ---EITLSEIKRDISNLQTLSSHQSDQLKEKLFRIKQLQVNNRRNDEEKLVSLRF 888
E+ L E K D+S L + +S S E++L + + ND+ K + L
Sbjct: 840 LYTELALKEAKEDLSFLTSEMSSSTSG-----EKLSEAKKHKINDKTKTTLIA 890

Query: 889 ELEDCEAALLDLAASLAKESQKNKSLIRQCAQL---ESQCBQLSQQLMIFSRQLESDYQ 944
D L + +E ++ + L +Q+ L E + ++ + L L+E Y

-452-

5
 10
 15

Sbjct: 891 LRDRQRIKLGHLGTYTERELKEMKRLYKQKTTLLKDESVKLGRMEVELDNLQLYLRYS 950
 Query: 945 MTLDEAKVGVANVLEILMARQLKSLQAKLALGPVNDIAQFEEVHERLITFATQRDD 1004
 ++ A K K + D AR++K ++ I+ LG VN+ +I +FE V+ER FL+ Q++
 Sbjct: 951 LSFEGAKERYQLETDPEARKGVKLKLALELTQVNLGSDIEFERNVRYKFLSQEKD 1010
 Query: 1005 LVHARKNLLELTITTDNDQVTKRFTFRAIRHSFETVQMRGGSSADLLITE-GOLLSA 1063
 L AKN L + I +MD+E+ RF TF IR F + F +FGGS A+L L+ DL+
 Sbjct: 1011 LTEAKNTLQVIEEMDEETKRFNDTFVQIRSHFDQVPSLFGGSRALRLTDPNDLH 1070
 Query: 1064 GVDIVSQPGKKIQSIALNMSGGRKALSALALLFAIRVKTIPVILDEVEALDEANVR 1123
 GV+I QPGKK+Q+LAL+SGE+AL+A+ALLF+I++V++ +PF +LDEVEALDEANVR R
 Sbjct: 1071 GVTIQAQPGKKLQNLNLSGGRKALTATALLFSILKVRPVFPVCLDEVEALDEANVR 1130
 Query: 1124 FQDYLARFDKSSQFIVVTHRGKMSAADSIVGVTMGESGVSKIVSVLKEAQE 1176
 F YL ++ +QFIV+THRGTM AD +YGVTMQESGVSK++SVL+E +E
 Sbjct: 1131 FAQYLKXYSSTQFIVVTHRGKMEADVLYGVTMQESGVSKIVSVLKEATKE 1183

20

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1163> which encodes the amino acid sequence <SEQ ID 1164>. Analysis of this protein sequence reveals the following:

possible site: 15

25
 30

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -4.99 Transmembrane 1092 -1108 (1088 -1110)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2996 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

35
 40
 45
 50
 55
 60
 65

>GP:CAB13467.GB:Z9112 chromosome segregation SMC protein homolog
 [Bacillus subtilis]
 Identities = 441/1192 (36%), Positives = 729/1192 (60%), Gaps = 25/1192 (2%)
 Query: 1 MFLKELELGGFKSFADKTIKIPDGVTAVVGPNGSGKSNITELRWALGESANNLRGGK 60
 MFLK +++ GFKSFA++ ++F KGVTAVVGPNGSGKSNIT+++RW LGE SA++LRGGK
 Sbjct: 1 MFLKRLDVIGFKSFAERISVDVFVGVTAVVGPNGSGKSNITDAIRWVLGQSARSLRGGK 60
 Query: 61 MPDVIIPAGTONRNPLNKAIVAVVLNDSHFIFTAKKEIRVERHIYRNGDSVYLDGRKVR 120
 M D+IPAG+ +R LN A+V +LDN DHF+ E+ V R +YR+G+S+LI+ + R
 Sbjct: 61 MEDIIIPAGDSRKRNLAEVITLNDNDHFLPIDFHEVSVTRVYRGSSEFLINMQPCR 120
 Query: 121 LKDIDLFDMDTGLGRDSFSIISGKVEIFNSKPEERRAIFEEAAGVLKYTKKGTQIK 180
 L+DI DLFD+GLG+++FSIISG+VEEI +SK E+RR+IFEEAAGVLKYTKRIG+ + K
 Sbjct: 121 LKDIDLFDMSGLGKEAFSIISGKVEILSSKAERRSIFEEAAGVLKYTKKGAENK 180
 Query: 181 INQTQDNLRLEDIITYELDTQALPLEKQAKVAKQFLELDANRQQLDILVKDIDLAQER 240
 L + TQDNL+R+EDI++EL+ Q+ PL+ QA +AK +LE + +++ + DI+ +
 Sbjct: 181 LFETQDNLRNVEDIINHELGQVSPKIQASIAKDYLEKKKELEHVEIALTYDIEKLGK 240
 Query: 241 QTKDTEALANQODIASTYAKRQSMREDDYQKFKQKQVLSQSDQTOTTLRLK/LK/LADL 300
 + E+ + + + + + E+ + + K Q L + + + Q L L + + L
 Sbjct: 241 WSTLKEKQVMAKERRLEASSAKSAKAKIDTRDKIQALDESVMELQVILATVSEKLEK 300
 Query: 301 EKQILVLKLSQGRARKKAKAKHLLQQLQQLDGFQARRKQCTQLH-----IDQQ 353
 E + E+K E+K A + BQL+E + PQ +E E+L + ++
 Sbjct: 301 EGRKGVLK-----ERKKNVQNBQLREAVIQPQKQETVLKRELSQRAVFETLQARV 353
 Query: 354 CDVKQQLNELSNALERFSSDPQIMSTLREEFVLIMQKRAALSQ/LALKAHLDKQKAR 413
 ++ Q+ E NL + + + + +E L + + L+ +A++ N+L L + +
 Sbjct: 354 KQLSAQVKEKQQLSNLENENVEETLQKLSDYFELNLSQASITNELQLLDQMQSQAVAL 413
 Query: 414 QHKAQFQLLIATKLQDNDESQKAQHYKAKQKQVHMLQNYQEGDKRQVQLERDYQLNQ 473
 Q A + + + + + + + + + + + Y++ + + + +R Y+ N+

-453-

Sbjct:	414 QRLADNRREKHLQEHSDISARKAACETEFARIQGEIHSQVGYRDMQTKYEQKRQYEQMB 473
Query:	474 ERLFDLLDQKKKEARKASLESIQKSHSQFYGVAVRLQSQKKIGITIGAVSEHLSFSD 533
5	L+ + + + + LE++Q S FY GV+ VL++++LQGI GAV E + S +
Sbjct:	474 SALYQAYQYVQOARSKKMLSTMGQFSQFYQGVKEVKKAKERLGGINGAVIALISTEQK 533
Query:	534 YQTALFVALGANSQHIIVTDEAAAKRAIAYKQROGRATPLPLTITKARSLSEHYHQRL 593
Y+T+E+ALG++Q++ DE +A++AI YLK+N GRATPLP+ I+ R L	
Sbjct:	534 YETATIALGASAGHVVTDDQSARKATQYKQNSPGRATPLPLSVIRDRQLQSRDAETA 593
10	Query: 594 ATCEYGLTFAESLIRYDQSLALIQNLSSATPFTIDQNTAARLGLYKRVIVLQGT 653
A +LG A L+ D+ + +IQNL+ I E + AN A+LG+ RIVL+G	
Sbjct:	594 ARHSSFLGVASELVTFDPAYRSVIQNLGLTFLITEDIKAGNELAKLGHRYRIVTLKSDV 653
15	Query: 654 LKPGGSPGGANQSNTPPI--KPELQISREI/TRIVQQLKITEKKEVALQSLIAKKE 711
+ PGG +GA ++ N + + ELE ++ L + E+ + E+V L+ + + +	
Sbjct:	654 VNPGGSMGGAVKKNNSLLGRSKLELDVTKRIASMEKTALEQEVKTLKHSIQMEK 713
20	Query: 712 LTQLKLAGDQARLQ--RAGMAYQQLQEQHDSKALLAALQSQSTTHDESILAPQARI 769
L L+ G+ FL +Q + Q+ Q+ EK ++ L ++E + SDE ++ +	
Sbjct:	712 LADLRETGEGLRLQDQVQOQYELQVAKNNINHLLEYDQKSALESDEERKVRKRL 773
Query:	770 ZEALTAIAKKKIALTCDDIDKKNKILIQKQNTIHOALQARLQERDLNNSKEQFQANQ 829
EE L+A++K L DID + + K + + + L++ + + K E+ N	
25	Sbjct: 774 EEELSAVSEKQKLEEDIDRLTKQKQSQSTKESLNEZELKIAAAKKEQACKSEKDL 833
Query:	830 SRLRTQLAQCCQNLKLESILNNNSQDSIQRLPQWQQLQCATHEKSGAQKRLVQGRFE 889
+RL+ L+ + + + + L+ S+ + +L+A+HK + + + L	
Sbjct:	834 ARLKRIELTETELAKKEKEDLSFLTSEMSS--TSGEKLEEAHKHKINDKTTIELIAL 891
30	Query: 890 IRDYEARLEETAEKITESEKNDTPIRRQTKL---ETHLEQVANRLRAYAKSLSEDFQM 945
D +L+ + + +E ++ + + + L E L + + L + L E + +	
Sbjct:	892 RRDQRILQHGLDITYERELKEMKRLYKQKTLTKLDEVKLGRMEVELNLQYVREYSL 951
35	Query: 946 TLADAKEVINSIDHLESAKEKLHLKQTLRALPQINSDAQNYEEVHERLFTLSQKTDL 1005
+ AKE E A++++ ++ I LG +N +I+++ E V+ER FL+ QK DL	
Sbjct:	952 SPFGAKERYQLETDPEAKARKVKLIKLAIEGLTVNLGSDIEFERVNERKFLSEQKEDL 1011
40	Query: 1006 TKAQNLLETINSMDSEVIGARPKVTFEAIQKSPKFTFTQMPGGGSAIDLVLTE--TLLSAG 1064
T+AKN L + I MD E+ RF TF I+ F + F +PGG A+L LT+ DLL +G	
Sbjct:	1012 TEAKNTLPQVIEEMDEMTKRFNDTFFVIRSHFDQVRSFLGGGRLRLTDPNLHLSG 1071
Query:	1065 IEISVQPPGKKIQSLNLSGGEGKALALALFAIRVKTIPFVLDEVEALDEANVRFP 1124
+EI QPPGKK+Q+INL+SGE+AL+A+LLF+I++V+ +PF +DEVEALDEANVRFP	
45	Sbjct: 1072 VELIAQPPGKKIQSLNLSGGEGALTAIALLSILKVRFPVFCVLEDEVEALDEANVRFP 1131
Query:	1125 GDLNRFKDKSQFIVTVTHRGTMADSIYGIYTMQESGVSKIVSVLKEAQ 1176
+L ++ D+QFIV+THRGTM AD +YH+TMQESGVSK++SVKLE +E	
50	Sbjct: 1132 AQVLAKYSSDTQFIVTVTHRGTMEDADVLYGVTHMQESGVSKIVSVLETK 1163

An alignment of the GAS and GBS proteins is shown below:

Identities = 732/1179 (62%), Positives = 911/1179 (77%)

Query:	1 MFLKRIEMQGGKSPADKTKVEFDQGVTAVVGPNMGSGKSNITSLRWALGESSAKSLGGK 60
Sbjct:	1 MFLKRIE+GPKSPADKTK+EFD+GVTAVVGPNMGSGKSNITSLRWALGESSAK+LGGK 60
Query:	61 MPDVIAGTENRKPLNYAQVSVTLNDSDFIRMTADEVRVERIFRNGDSKYLIDGRKVR 120
Sbjct:	61 MPDVIAGT+NR PLNYA+V+V LNSDHF+ E+RVER I+RNGDS+YLIDGRKVR 120
Query:	121 LRDIHDLFMDTGLGRDPSFISIQGRVEAIPNSKPEERRAIFEEAGVLKYTKRKETQSK 180
Sbjct:	121 LRDIHDLFMDTGLGRDPSFISIQGRVEEIPNSKPEERRAIFEEAGVLKYTKRKETQSK 180
65	Query: 181 LEOQGNLDRLIEDIYELMDQVQPLEKQASIAKRFVLDSERQGLHLSILEDILQFQSD 240
L QTK NLDRLIEDIYELD Q+ PLEKQA +AK+FL LD R+ L L IL+D+ Q	

Sbjct: 181 LNQTDQNLDRLEDIIYELOTQAPLEKQAKVAKQFLELDANRQQLQDILVQKIDIAQER 240
 Query: 241 LTTVEKLLTVRKELATTYYQQRQSLSDENQSLKQKNHLSSEIISAKOLALLDVTKASDL 300
 T E L +++LA+YK +RQS+E++ Q KQK+ LS+E + Q LL++TKL +DL
 5 Sbjct: 241 QTKTETALAAQQDLASTAYAKRQSMEDYQKFKQKQVLSQESDQTTTLELTKLIADL 300
 Query: 301 HRQDLIRLSSNQWAKKKKPAQRIARLIRKANDCSDOIYQKMIELTTSEKTAQTRSEI 360
 E+QI+L++LES Q+AKK BA + L +L+ + + Q +L + +++ ++
 10 Sbjct: 301 EKQTELVKLESQKARKKAKKHLRQ+QEQLDGFQAKKKQCTEQHLHDQQLCDVKQQL 360
 Query: 361 VSTESSLERFSPNDQILIKLRDPVTAQKEADTSSALTAIADIENQKQASQAKSQRI 420
 ++LEFS++FDQ++E LRB+PV LMQ+BA SN LTAL A ++ +QA Q K+QE
 Sbjct: 361 NELSHALERFSSDDQLMFTLRKEFVLIAQKGAALSNQITALKAHLLKCKQARHQKQAY 420
 Query: 421 QEVSNLEVLKSNAKVALEPFPAKKNVQRLLSHYQDLQQTQLNLECEYKMQSILFDHL 480
 Q + L+ L ++ A ++A K+ V LL +YQ+ + +Q LE +Y+ Q LFD L
 15 Sbjct: 421 QLLVTKLDQNDSSQAKQAHYKQKQEVMLQHYYQEDKRVQELERUYQIAQERLFDL 480
 Query: 481 DETKSKQARISLESILNKHNSFYAGVKSVLQAKDLOGIIGANSEHLSFDKHVYQTALEI 540
 D+ K K+AR +SLEST K+HS FYAG++VLQ++ +LGGIIGANSEHLEFD VQTALEI
 Sbjct: 481 DQKKGEARKASLESTQKSHSQFYAGVRAVLQSQKLGIIIGANSEHLSFDUYQTALEI 540
 Query: 541 ALGSSQHIIVDESAKRSIAPLAKNNRQGRATFLEPLTTIKPRELAQHVLSKQSSQGL 600
 ALG +SQHIIV DE+AKR+IA+LKNRQGRATFLEPLTTIK R L++HY +L + +Q+L
 25 Sbjct: 541 ALGANSHIIVDEAANKRAIYALKNNRQGRATFLEPLTTIKARLSSEHRYQATCEGYL 600
 Query: 601 GIASELVYDQRLSNIPKNNLGLTAIFDTVDNANVARQINQVRALTLDGTELPGSGY 660
 G A L+ YD LS I +N L TAI F+T+D AN+AR L Y+VR+VTLDTGTELPGSG+
 30 Sbjct: 601 GIASLIRYDQSLSAIIONLSSTAIPETIDQANIAARLLGKVRIVTLDTGTELPGSGF 660
 Query: 661 SGGANRQNTVFIKPELDNLKELKQACSKQLIQKEVAILLSQLKEKQETLAQIKQDGE 720
 SGGANRQ+NT FIKPEL+ + +EL + + I EKEA L L K+E L QLK G+
 Sbjct: 661 SGGANRQNTVFIKPELQISBELTRLVBQLKITEKVAALQSDLIANKBELQQLKAGD 720
 Query: 721 QARLEQRADIEVQQLSEKLADNLKYLNGQLSSGALEQTTSENEKNRLEKELEQPAIKK 780
 QARL BORA + YQQL EK D L L L S + E+ R+E L A K K
 35 Sbjct: 721 QARLABQRAQMAVQQLKEQKQSDKALLAALDQSQTTTHSDSLABQARTEALTAIAKK 780
 Query: 781 BELTTSIAQIKEDKDSIQEKNVNLITLLSBAQLEERDLNQEKFERNANCTRLBILTSEIK 840
 LT I IKR+KD I++K N+ LS+A+L+ERDLANE+KFE+AN +RL L + +
 40 Sbjct: 781 NALTCDIDDIKENKIDKQNTIHQALSQARLQERDLAKKFEQANQSRLETQLKQKQ 840
 Query: 841 RDISNLQTLSSQDSQLDKEELPRIKQLQVNRNREMBEKLVSIRFELEDCEAALDDL 900
 ++I L++L++ SQ + LQ+ +KQL + +++IN LRFE+ED EA L++
 45 Sbjct: 841 QNLIKLESILNNVQSDSQRLPQWQQLQDATEHKSQAQKRLVQLRFEIDYARLAEET 900
 Query: 901 AASLAKEQKQNSLRQQAQLESQCBQLSQQLMIFSRQLSEDPQMTLDEAKVIANVLEDI 960
 A + KE +KN++ IR+Q +LR+ RQ++ +L +++ LSED+QMTL +AK N ++ +
 50 Sbjct: 901 AEKITKESEKNDTFRRTQKLETHLEQVANRLRAYAKLSLEDQPMTLAKAEVTVMSIDEL 960
 Query: 961 LGAREQLKSLQAKIQLSPVNIDATAQFREVHERLTFIANQREDLVHAKNNLLETITMD 1020
 A+E+L LQ I+ALGP+N DAI Q+ERVHERLTFI+Q+ DL AKNNLLETI MD
 Sbjct: 961 ESAKEKLHLQKTRALGPINSDAINQYEVHERLTFITSQCTDLTKAKNNLLETINSD 1020
 Query: 1021 DEVKTRFKSTFEAIRHSFKETFPQMFGGSDLLIYEGDILLSAGVDSIQVPPKKIQSLN 1080
 EVK RPK TFEAL+ SFKETP QNFGGGSADL+LIE DILLSAG++ISVQPPKKIQSLN
 55 Sbjct: 1021 SEVKARFKVTFEALQKSFKEFTPTQMFGGSDLVLTETDLLSAGITRISVQPPKKIQSLN 1080
 Query: 1081 LMSGGEKALSALALLFAIRVKTI PFVILDEVBALDEANVRKFGDYLNRPKDSQPIVV 1140
 LMSGGEKALSALALLFAIRVKTI PFVILDEVBALDEANVRKFGD+LNRPK SQPIVV
 60 Sbjct: 1081 LMSGGEKALSALALLFAIRVKTI PFVILDEVBALDEANVRKFGD+LNRPKDSQPIVV 1140
 Query: 1141 THRKGTMGAADSISYITMQESGVSKIYSVKLKEAQEMTN 1179
 THRKGTM+AAISYIG+TMQESGVSKIYSVKLKEAQEMTN
 65 Sbjct: 1141 THRKGTMGAADSISYITMQESGVSKIYSVKLKEAQEMTN 1179

SEQ ID 1162 (GBS199) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 2; MW 75kDa).

GBS199-GST was purified as shown in Figure 208, lane 3.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 359

A DNA sequence (GBSx0390) was identified in *S. agalactiae* <SEQ ID 1165> which encodes the amino acid sequence <SEQ ID 1166>. This protein is predicted to be ribonuclease III (mc). Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3372 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9711> which encodes amino acid sequence <SEQ ID 9712> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB13466 GB:299112 ribonuclease III [Bacillus subtilis]
Identities = 115/230 (50%), Positives = 154/230 (66%), Gaps = 1/230 (0%)

Query: 13 KKKELRSKLEKDYGVFANQELLDTAFTHTSYANEHRLNISHNERLEFLGDAVLQLLI 72
          KK+++ + E+ + F N++LL AFTH+SY NEHR NERLEFLGDAVL+L I
Sbjct: 15 KKVQEQKEPQER- ISVHFQNEKLLYQAFTHSSVYNEHRKPKPYEDNERLEFLGDAVLELTI 73

Query: 73 SQYLSTKYPQKAEGDLSKLRSMTIVREESLAGFSRLCGFDHYIKLGKGEESGGRNRDITL 132
          S++LF KYP +EGDL+KLR+ IV E SL + F + LGHSEE +GGR R +L
Sbjct: 74 SRFLPAKYFAMSEGLTKLRRAAIVCEPSLVSLAHELSTFGDLVLLGKGSEMTGGRKRPAII 133

Query: 133 GDLFSAFLGALLLDKGVVHVAFVNMVMIPIHVKGTYYERVIDYKTSLQELLQSHGDVKID 192
          D+FEAF+GAL LD+G+H V +F+ + F + G + V D+K+ LQE +Q G ++
Sbjct: 134 ADVFSAFAGALYLDQGLRPVRSFLKVVYVFVKINDGAFSHVMDPKSLQEQYVQRDGRKGLSE 193

Query: 193 YQVTNBSGPAHAKPEFTVSVNQRTNSQGIGRSKAAEQDAAKNALATLQ 242
          Y++NE GPAAH +EFE VS+ E L G GRSKK AEQ AA+ AIA LQ
Sbjct: 194 YKISNRKGPAAHNRFEATVSLKGRPLGVNGRSKCAEQHAAQALAKLQ 243
```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1167> which encodes the amino acid sequence <SEQ ID 1168>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1414 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

Identities = 170/227 (74%), Positives = 192/227 (83%)

```
Query: 15 MKELRSKLEKDYGVFANQELLDTAFTHTSYANEHRLNISHNERLEFLGDAVLQLLIS 74
```

-456-

```

      MK+L  L  + I F +  LL+TAFTHTSYANEHRLIN+SHNERLEFLGDAVLQL+IS+
Sbjct: 1  MKQEELLSTSTPDIQNDLILLETAPTHTSYANEHRLINSHNERLEFLGDAVLQLIIS 60

Query: 75  YLFTKYQKAGDLSEKLSMTVRESLAGPSRLGFDHVIKLGKGEKSGGRRDITLGD 134
      YLF KYF+K  EGD+SKLSMTVRESLAGPSR C FD YIKGKGEKSGGRRDITLGD
Sbjct: 61  YLFAKYFKKTEGCMSEKLSMTVRESLAGPSRFGSDAVIKLGKGEKSGGRRDITLGD 130

Query: 135  LFEAPFLGALLLDKRGVVFVHAFVNKVMIPVVEKGYTERVIDYKTSLSQLQLQSHGVKIDYQ 194
      LFEAPFLGALLLDKRG++ V F+ +VMIP VESK +ERVVDYKT LQE LQ+ GGV IDYQ
Sbjct: 121  LFEAPFLGALLLDKSIDAVRFVFLQVMIPQVRKGFERNVDYKTCLEQLQKGVKIDYQ 190

Query: 195  VINESGPAHAKGFVETVSVFQNLGSGGIGRSKGAABQDAANALATT 241
      V +E GEAHAK+FEV++ VN  LS+G+G+SKK ARQDAANALA L
Sbjct: 161  VISSKGPAPAKQFEVSVVNGAVLSEKGLGSKKLABEQDAANALALQ 227

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 360

A DNA sequence (GBSx0391) was identified in *S. galactiae* <SEQ ID 1169> which encodes the amino acid sequence <SEQ ID 1170>. Analysis of this protein sequence reveals the following:

```

Possible site: 43
>>> Seems to have a cleavable N-term signal seq.
      INTEGRAL Likelihood = -4.19 Transmembrane 100 - 116 ( 99 - 117)
      INTEGRAL Likelihood = -2.44 Transmembrane 81 - 97 ( 81 - 97)

----- Final Results -----
      bacterial membrane --- Certainty=0.2678(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:CAC12789 GB:AJ279090 hypothetical protein [Staphylococcus
      catenous]
      Identities = 50/114 (43%), Positives = 72/114 (62%)

Query: 3  KIFYISLGFSLGIGIAGIVLPVPTPLVLLSAPCFSSSEKFDIKLRQTKVYKYAAD 62
      K ++LG I GIG GIV+P++PTTP +LL+A CFSRSS+K+ ML TK++ Y
Sbjct: 2  KYVLMTLGILFAGIGFVGI VELLPTTFPLLLAACPSRSSSKFKRKLWTKIHDEYVS 61

Query: 63  FVRSRIAPARKKSMIWQIVILMGISIVYFAPLMWLKGLLIGTIVGTIVLFTVV 116
      F + ++K ++ +YILMGISI+ +++++ LLI V T VLF V
Sbjct: 62  FKRDKGPTLKKKPKLLTSLYILMGISIFITENLYIRITLLIMLFVQTVLFTFV 115

```

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 361

A DNA sequence (GBSx0392) was identified in *S. galactiae* <SEQ ID 1171> which encodes the amino acid sequence <SEQ ID 1172>. Analysis of this protein sequence reveals the following:

```

Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1908(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1173> which encodes the amino acid sequence <SEQ ID 1174>. Analysis of this protein sequence reveals the following:

Possible site: 45

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1610 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 225/269 (83%), Positives = 246/269 (91%)

Query: 1 MSEIGFKYSILASGSGTNCIFYETPQKRLIDAGLTGKKVSLAEINRKPOLDAILVT 60
M+E GFKYSILASGSGTNCIFYETP+KRLIDAGLTGKK+ISLAEI+RKPEDLDAIL+T
Sbjct: 1 MMSGFKYSILASGSGTNCIFYETPVKGLIDAGLTGKATISLAEIDRKPOLDAILIT 60

Query: 61 HEHSDHIKGVGVILARKYHLDIYANEQTVKVMDERNMLGKVDVSQRHVFGKKTITFGDLD 120
HEHSDHIKGVGV+ARKYHLDIYANE+TW+MDE NMLGK+D SQKH+P R K LITFG+D
Sbjct: 61 HEHSDHIKGVGVNARKYHLDIYANEKTWQMDENMLGKLDASQKHIFQDKVLTFGDVD 120

Query: 121 IESFGVSHDAVDPOQFYRNMKDKSFVMLTDTGYVSDRMAGLIENADGYLIESNHDIELR 180
IESFGVSHDA+DPQFYR+MKD+KSFVMLTDTGYVSDRM G+IENADGYLIESNHDIELR
Sbjct: 121 IESFGVSHDAIDPOQFYRINKDKSFVMLTDTGYVSDRMGTGIENADGYLIESNHDIELR 180

Query: 181 SCSYPWTLKQRIKLSDLGHLNSNEDGSETWIRTIGNRTKHYLGHLSKENNIKSLAHWTM N 240
SCSYPW+LKORILESD GHLNSNEDG+ MTR+GG TK IYLGHLSKENNIKSLAHWTM N
Sbjct: 181 SCSYPWNLKQRIKLSDLGHLNSNEDGAGAMIRSLGYNTKHYLGHLSKENNIKSLAHWTM N 240

Query: 241 NLMRADPVGUGTDFSVHDTSPDGAFLTRI 269

L AD VGTDF+VHDTSPD+A PLT I

Sbjct: 241 QLAMADLAVGTDFTVHDTSPDGAFLPTDI 269

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 362

A DNA sequence (GBSx0393) was identified in *S.agalactiae* <SEQ ID 1175> which encodes the amino acid sequence <SEQ ID 1176>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have an uncleavable N-term signal seq

INTERAL Likelihood =-11.94 Transmembrane 15 - 31 (5 - 34)

----- Final Results -----

bacterial membrane --- Certainty=0.5776 (Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1177> which encodes the amino acid sequence <SEQ ID 1178>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 335/443 (75%), Positives = 392/443 (87%)

```

5   Query: 7   NIRSFLALLFLVFAVYFVYLAVRDFKMSKNIKLNLNKKVRDLIAGNYSILIQGDAD 66
      N+ +SELA+L LLVFA YP+LAVRD++ ++ IR+++ K+RDLI G Y+d I + D +
      Sbjct: 8   NLSTFELAILILLVFAVYFIHLAVRDYRNARIIMMSHKIRDLINGRYTDLIDSKADIE 67

10  Query: 67   LVELGSDINDLSDVFRMAHNDLQEKKNRLASIIITMIDGVIATDRSGKIVIMINETAQQQF 126
      L+EL + LNDLSDVFR+ H+NL QEKKNRLASII YH+DGVLAIDRSGKI+MINETA++Q
      Sbjct: 68   LVELSDQINDLSDVFRILTHKNLAQRKNRLASIIAYMSDGVLAIDRSGKIIIMINETAQQL 127

15  Query: 127  NLAYDEALSMNIVMISGSGSPYSPQDLVSKTPEVVLARRDENQEFVTLRIRPALNRRESG 186
      NL+ +EAL NI D+L + Y+++DLVSKTP V +N R++ GEFV+LR+RPAI NRRESG
      Sbjct: 128  NLSKEBALKGNITDLLEGDTSYTYRDLVSKTPVVTVNSRNDMGEFVSLRLRPAI NRRESG 187

20  Query: 187  FISGLVAVSHDANTEQKEKERERRLFVSNVSHELRTPLTSVKSYLEALDEGALNEVAPSF 246
      FISGLV V HD TEQKEKERERRLFVSNVSHELRTPLTSVKSYLEALDRGAL E++APSF
      Sbjct: 188  FISGLVVVLHDTTEQKEKERERRLFVSNVSHELRTPLTSVKSYLEALDEGALKEDIAPSF 247

25  Query: 247  IKVSLCETNRMMRMISDLLSLSRIDNEVTHLDVEMTNFTAPNTSIINRFPDQIRNQKTVTG 306
      IKVSLCETNRMMRMISDLL+LSRIDN+VT L VEMTNFTAP+TSLINRFPD ++NQ T TG
      Sbjct: 248  IKVSLCETNRMMRMISDLLSLSRIDNQVTLAVEMTNFTAPNTSIINRFPDLVGNQKTGTG 307

30  Query: 307  KYVEIVRDYPLKSNVEIDTDMQTVQIDNIIINWAKVYSPDGGKITVNLRTKTMILISIS 366
      KYVEIVRDY+P S+N+EID DMKQVI+NIILN++KYS PDGGKITV ++T Q+I+SIS
      Sbjct: 308  KYVEIVRDY+PITSVNIEIDNDKMTQVIRNIIINWAKVYSPDGGKITVNMKTDTQLIISIS 367

35  Query: 367  DQSLGIPKNDLPLIFRPFYVDKARSRQGGTGLGLSIAKEIVKQHNFWIWAKESEYKGS 426
      DQSLGIPK DLPLIFRPFYVDKARSR QGGTGLGLSIAKEI+KQH GPFWAKS+YKGS
      Sbjct: 368  DQSLGIPKNDLPLIFRPFYVDKARSRQGGTGLGLSIAKEI+KQHGPFWAKS+YKGS 427

      Query: 427  TFFIVLPYDKDAVTEENEDVED 449
      TFFIVLPY+KDA YEEN+ D
      Sbjct: 428  TFFIVLPYERDAAYEENEDVD 450
  
```

A related GBS gene <SEQ ID 8561> and protein <SEQ ID 8562> were also identified. Analysis of this protein sequence reveals the following:

```

40  Lipop: Possible site: -1   Crend: 8
      MCG: Discrim Score:      8.59
      GVH: Signal Score (-7.5): -3.38
           Possible site: 26
      >>> Seems to have an uncleavable N-term signal seq
      ALOM program count: 1 value: -11.94 threshold: 0.0
45  INTEGRAL Likelihood = -11.94   Transmembrane 15 - 31 { 5 - 34)
      PERIPHERAL Likelihood = 8.27   178
      modified ALOM score: 2.89

      *** Reasoning Step: 3

50  ----- Final Results -----
           bacterial membrane --- Certainty=0.5776(Affirmative) < succ>
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
55
  
```

The protein has homology with the following sequences in the databases:

```

67.5/83.5% over 439aa
      Streptococcus pneumoniae
      GP[5830524] histidine kinase Insert characterized
60  ORF01458 [331 - 1647 of 1947)
      GP[5830524] [emb] CAB54569.2 || AJ006392 (10 - 449 of 449) histidine kinase {Streptococcus
      pneumoniae}
      %Match = 45.6
  
```

```
%Identity = 67.5  %Similarity = 83.4
Matches = 297  Mismatches = 70  Conservative Sub.s = 70
```

[illegible]

SEQ ID 1176 (GBS41) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 4 (lane 7; MW 50kDa), in Figure 168 (lane 2-4; MW 65kDa – thioredoxin fusion) and in Figure 238 (lane 4; MW 65kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 13 (lane 7; MW 75kDa).

50 Purified Thio-GBS41-His is shown in Figure 244, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 363

A DNA sequence (GBSx0394) was identified in *S.galactiae* <SEQ ID 1179> which encodes the amino acid sequence <SEQ ID 1180>. This protein is predicted to be VicR protein (regX3). Analysis of this protein sequence reveals the following:

Possible site: 60
>>> Seems to have no N-terminal signal sequence

-460-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2754 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1181> which encodes the amino acid sequence <SEQ ID 1182>. Analysis of this protein sequence reveals the following:

Possible site: 60

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2754 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 205/236 (86%), Positives = 221/236 (92%)

Query: 1 MKKILIVDDKPISDIIFKPNLTKEGISTATAFDGREALVQYARPOFDLIILDLMLPELDG 60
 Sbjct: 1 MKKILIVDDKPISDIIFKPNLTKEGIDIVIAFDGREAVIFPSEKPDLIILDLMLPELDG 60

Query: 61 LEVAKEVRKTSHPFIIMLSAKDSEFDKVGLEIGADDIVTKPFSNRRELLARVKAHLRRT 120
 Sbjct: 61 LEVAKE+RTSH+FIIMLSAKDSEFDKVGLEIGADDIVTKPFSNRRELLARVKAHLRRT 120

Query: 121 NIETAVAEASQNASDDITIGELQILPDPAFAKVRGSEIELTHREFELLHHLATHIQVM 180
 Sbjct: 121 NIETAVAE+A++ +TIG LQILPDPAFAKK Q+P+RLTHREFELLHHLA H+QVM 180

Query: 181 TREHLLTVMGYDYPGDVRIVDVTVRLREKIEDTFGRFEYILTRGQGYMKSY 236
 Sbjct: 181 TREHLLTVMGYDYPGDVRIVDVTVRLREKIEDTFGRFEYILTRGQGYMKSY 236

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 364

A DNA sequence (GBSx0395) was identified in *S.agalactiae* <SEQ ID 1183> which encodes the amino acid sequence <SEQ ID 1184>. This protein is predicted to be amino acid ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3791 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CB14701 GB:Z99118 glutamine ABC transporter (ATP-binding protein) [Bacillus subtilis]
 Identities = 149/244 (61%), Positives = 200/244 (81%), Gaps = 2/244 (0%)

Query: 3 LISYKRVNKKYGDYHALRQINLEI RFGQVVVLGPGSGSKSTILIRTNWALESIDDGLV 62
 I+++NVNK+YGD+H L+QINL+IE G+VVV++GPGSGSKSTL+R +N LESI++G L V
 Sbjct: 1 MITPCNVNKHGYDHFVLKQINLQIRKQGVVVVIGPGSGSKSTILIRCNLESINRGVLT 60

-461-

Query: 63 NGHELANTSSKELNLRKEVGMVFOHFLYPHKTVLENITLAPIKVLQSKKRAMBIAEK 122
 NG + N ++ ++ +GMVFGHF+LYPHKTVL+NI LAP+KVL+QS ++A E A
 Sbjct: 61 NGTAT-NDKRTDINQRCNIGMVFOHFLYPHKTVLQNIPLAFVVLRLQSPBQAKETARY 119

5 Query: 123 YLKFVNMWERKDSYPMGLSGGQKQRIATARGLAMHPKLLLFDEPTSLDPETIGDVLV 182
 YL+ V + + +YPS LSGGQ+QR+AIARGLAM P++L+LDEPTSLDPE IG+VL VM
 Sbjct: 120 YLEKVGIPDKADATPSQLSGGQQQRIATARGLAMKPEVMLFDEPTSLDPENIGEVLDVM 179

10 Query: 183 QKLANDGMNMVVVTHMGFAEVADRILIFWADGEILVDITDVOFFDNKPSRAKQFLSN 242
 + LA +GM MVVTHMGFA+EVADRI+P+ +G+IL + +P+ NP+S RA+ FLS
 Sbjct: 180 KTLAKGNTMVVVTHMGFAEVADRIVFDGKILESAPVA-EFTANPKSERARLFLSR 238

Query: 243 IINH 246
 I+NH
 15 Sbjct: 239 ILNH 242

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1185> which encodes the amino acid sequence <SEQ ID 1186>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3763 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 25 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 131/243 (53%), Positives = 179/243 (72%), Gaps = 2/243 (0%)

30 Query: 2 SLISYKNVKNKYGDVHARQINLEIEPGQVVLLGPSGSGKSTLIRTNVALESIDDGSLV 61
 ++IS K+++KYNG L+ I+L+I PG+VVV++GPSGSGKSTL+RTNM LE G +
 Sbjct: 5 AIIISIKDLHKYGENEVLKSIDLIMPGEVVVIGPSGSGKSTLLRTNVLLEVPKQQR 64

35 Query: 62 VNGHELANTSSKELNLRKEVGMVFOHFLYPHKTVLENITLAPIKVLQSKKRAMBIAEK 121
 G ++ + ++ ++R++GMVFO FNL+P+ T+LENITL+PIK +K EA + A
 Sbjct: 65 FEGIDITD-KKIDIFGRKNGVMVFOQFNLFPNMTILENITLSPIKTKGMAKADKITAL 123

Query: 122 KYLKFVNMWERKDSYPMGLSGGQKQRIATARGLAMHPKLLLFDEPTSLDPETIGDVLV 181
 L V + E+ +Y+ LSGGQ+QRIATARGLAM P +LLFDEPTSLDPE +G+VL+V
 40 Sbjct: 124 SLIDKVLSEKAKAYPASLSGGQKQRIATARGLAMDEVLVLFDEPTSLDPENIGEVLA 183

Query: 182 MQKLANDGMNMVVVTHMGFAEVADRILIFWADGEILVDITDVOFFDNKPSRAKQFLS 241
 MQ LA GM NV+VTHMGFA+EVADR++PM DG ++V+ FD +E R K FLS
 45 Sbjct: 184 MQDLAKSGMTNVIVTHMGFAEVADRVMPM-DGGVIVEGSGINQLFDLTEREKTFLS 242

Query: 242 NII 244
 ++
 Sbjct: 243 RVL 245

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 365

A DNA sequence (GBSx0396) was identified in *S.agalactiae* <SEQ ID 1187> which encodes the amino acid sequence <SEQ ID 1188>. This protein is predicted to be glutamine-binding. Analysis of this protein
 55 sequence reveals the following:

Possible site: 27
 >>> Seems to have a cleavable N-term signal seq.
 ----- Final Results -----

-462-

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB73178 GB:AL139076 probable ABC-type amino-acid transporter
 periplasmic solute-binding protein [Campylobacter
 jejuni]

Identities = 99/240 (41%), Positives = 141/240 (58%), Gaps = 3/240 (1%)

10 Query: 1 MLARKRLTYFLLSCIFIFLLFYNSTANQLSEIKKSGVLKVGKQDVPMFGYNASTNQ 60
 M+ RK L + + + F + + + L IK G L VGVK DVP+ + + T +
 Sbjct: 1 MVRKSLLLKLAFLGACVAFSNANAEKGLSEIKSKQQLTVGVNDVPHYALLDQATGE 60

15 Query: 61 YEGMEIDIAKKIASKL--GKPVFVPTTAQTRFPLMDNQIDILIAITYTTPPERKANYN 117
 +G E+D+AK +AKS+ K V A+TR PL+DNG +D +IAT+TITPERK YN
 Sbjct: 61 IKGFEDVDVAKLAKSLGDDKKIKLVAVNKTRGFLDNKGSVDVIATPTTTPPERKIRYN 120

20 Query: 118 ISKAYYHDEIGFLVRNHSHTIKELDKGKHIGVAQATTVMNLEKYAKHKLFPSYALG 177
 S+ YY D IG LV K K+ + + G +IGVQV ATTK + + AK+ + + +
 Sbjct: 121 FSEPTVQDAIGLLVLKCKKYSGLADMGANIGVQAQATTTKAIGEAARKIGIDVKSFEFP 180

Query: 178 SFPELAISLYANRIDAFSVDKSIILSGYLSPHITTLKEGPNTOEYGIATSKQDKVLIPYIN 237
 +P + +L A R+DAFSVDKSIILGY+ +IL +F Q YGI T K D YN+
 Sbjct: 181 DYSIKAAALDAKRVDAFSVDKSIILGIVDDKSEILDFSEFQSGIVTKDDPAFAKYVD 240

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1189> which encodes the amino acid sequence <SEQ ID 1190>. Analysis of this protein sequence reveals the following:

Possible site: 30
 30 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -6.16 Transmembrane 17 - 33 (15 - 35)

----- Final Results -----
 bacterial membrane --- Certainty=0.3463 (Affirmative) < succ>
 35 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9097> which encodes the amino acid sequence <SEQ ID 9098>. Analysis of this protein sequence reveals the following:

40 >>> May be a lipoprotein

----- Final Results -----
 bacterial membrane --- Certainty= 0.000 (Not Clear) < succ>
 bacterial outside --- Certainty= 0.000 (Not Clear) < succ>
 45 bacterial cytoplasm --- Certainty= 0.000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 66/251 (26%), Positives = 111/251 (43%), Gaps = 27/251 (10%)

50 Query: 23 PNSTSANQLSELKSGVLKVGKQDVPMFGYNASTNQVGEIDIAKKIASKSLGVKPVF 82
 P+ + + IK+ GVLKV +YN +N+ G E+D+ K+I K L +K F
 Sbjct: 34 PHOSKSSWDITKEKGVLKVAITPQYPTQTSFYNDNNELWGYEVLVMEIKGIRLNLIKVKF 92

Query: 83 VPTTAQTRFPLMDNQIDILIAITYTTPPERKANYNISKAYYHDEIGFLVR-----KNSHIK 138
 V T +D+G+DI + + ITP+R+ YNIS Y + G +VR N K
 55 Sbjct: 93 VETGFDQAFSTVDSGRVDISLNNFDITPKRKKYINISTPYKVGVMGIVRADGSSNIK 152

Query: 139 TIKELDKGKHIGVAQATTVMNLEKYAKHKLFPSYALGSPPELAISLYANRI----- 191
 + + GK A G +K A+L + + + +Y N +
 60 Sbjct: 153 DLSDWGKKKAGASGTEYMKVAQKQ-----AELVTVYNDVINDVANGRTDF 203

Query: 192 --DAFSVDKSIILSGYLSPHITTLKE---GFMTQYGIATSKQDKVLIPYVKNILVSWRK 245

-463-

++ K + LS + + + +N E GI +K+D L ++ ++ K
 Sbjct: 204 IPNDYFAQKLFVDYHLSQNPILNVRMSDVQYNPTQSGIVWKKKIDSLKKIDAVIKMIK 263

Query: 246 DGSLLKHIVQKF 256

Sbjct: 264 DGSLLKISSEY 274

SEQ ID 1188 (GBS136) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 5; MW 29.9kDa).

The GBS136-His fusion product was purified (Figure 200, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 284), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 366

A DNA sequence (GBSx0397) was identified in *S. agalactiae* <SEQ ID 1191> which encodes the amino acid sequence <SEQ ID 1192>. This protein is predicted to be integral membrane. Analysis of this protein sequence reveals the following:

Possible site: 55
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -9.34 Transmembrane 32 - 48 (27 - 55)
 INTEGRAL Likelihood = -5.04 Transmembrane 200 - 216 (196 - 219)
 INTEGRAL Likelihood = -3.13 Transmembrane 93 - 109 (93 - 113)
 INTEGRAL Likelihood = -2.02 Transmembrane 74 - 90 (74 - 92)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4736(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CA873177 GB:AL139076 putative ABC-type amino-acid transporter
 permease protein [Campylobacter jejuni]
 Identities = 112/226 (49%), Positives = 160/226 (70%), Gaps = 3/226 (1%)
 Query: 5 NISPPAISRWGAFFMFDLPFKGFLYTLGISFALLIALILGLISGLSTSKSKVGLIS 64
 +ISPPA ++ ++ D F GE+YTL+S ALL+A I G + G ++TS+ K+ + +
 Sbjct: 25 SISPPAVWKFILDALNKDAFINQFITYLEVISILALLIATITPGTIGVWMTSRFKIIRAYT 84
 Query: 65 RIYVVFQNTPELLWQVVFVYVYGLAISNGHVMISAPFFAVLCVGLYHGAYISRVIRSGIR 124
 RIYV+QCN PL++Q+ F++Y L ++ + + F VL VG YHGRAY+SEV+RSQI
 Sbjct: 85 RIYVELQNVPLVIQIFFLFYALFVLG---IRLDIPTIGVLGVGAYHGAYVSEVRSGLI 141
 Query: 125 AVPGQCTEALAQGFATQNTMQLILILPQAVRTILPMTNQVNLKNTSTVALISGADIN 184
 AVP+QG EA+ +QGF O M+ II+PQ +R ILPMTNQ+VNLKNTS + I+ GA++M
 Sbjct: 142 AVPRGQFEASASQGFYTIQMQRYIIVPQTRIILPMTNQVNLKNTSVLILVGGAEIM 201
 Query: 185 FVAKAWAYDTTWYIFAPGAAIFYPVICPLASWARKQRELNKTY 230
 A ++A D NY EA+ AA+ YF+IC+FLA +A+ R KK +
 Sbjct: 202 HSADSYAADYGNAYAYIFAAVLYFIICPYLAFKAYENKLEKFAH 247

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1193> which encodes the amino acid sequence <SEQ ID 1194>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have a cleavable N-term signal seq.

-464-

INTEGRAL	Likelihood = -6.26	Transmembrane	307 - 323 (303 - 327)
INTEGRAL	Likelihood = -5.89	Transmembrane	485 - 501 (479 - 502)
INTEGRAL	Likelihood = -1.12	Transmembrane	375 - 391 (375 - 391)

5 ----- Final Results -----
 bacterial membrane --- Certainty=0.3506(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the databases:

>GF:BA17584 GB:D90907 glutamine-binding periplasmic protein
 [Synechocystis sp.]
 Identities = 146/532 (27%), Positives = 244/532 (45%), Gaps = 59/532 (11%)

15 Query: 6 YNKKLILSCIVALLALFGMSRAQANQYLRVGMENYAFNNVQDDASNGKAVPIGTSQY 65
 Y L L L L+A+A+ + Q + V E + FF T E T Q
 Sbjct: 16 YYLLALGVLLAIAIPLLPFASQVSRQTIIVATETFPFFPMTD-----ENTGQL 65

20 Query: 66 ANGVDVQVAKKAVAKMKELLVVKTSWTGLPALITSGKIDMIAAGMSPTKERRNISFEN 125
 G+DV + + +A + + + G+IPAL S + + + T ER +SYF+
 Sbjct: 66 T-GFDVDLIQATGEAAQVTVDIQGFDFDGIIPALQSNVTGAALSAITITPERAQSVFS 124

25 Query: 126 SSSYTSQFVLVVTANGKADATSLKDFSGAKVTAQQGVHVNLLTQLGAKLTQPMDFSQ 185
 + S VJ + +LKD G++ G + T +GAK+ T +
 Sbjct: 125 PYPK--VLAIAVQDNDTIGNKDLGKRLAVALGTGAMVATNVGAKV-TNFDITS 181

30 Query: 186 MRQALTSVIDAYISERPEAMTREAADSLKMLITLKKGFVAESDAALVGMKINDRVA 245
 Q L +G DA I+RP + A D+ L+ + + +E I+ + + +
 Sbjct: 182 ALQELVNGADAVINDRFLVLA- IKDGLNFKISADV-SEDYGIAMFLAPPG-- 236

35 Query: 246 TVNQVLESGSQTDHMLMDVMYTKQFVEKKAEDAKASFLQWAIIFGN----- 294
 +NQ E +Q +++++ EK + FL + K
 Sbjct: 237 -INQIREVLNQ-GLQFIENGTNATYKVGFGKNPPFLPLNAPSIVGVKGTAGSLTERS 294

40 Query: 295 -----WKQLAGTQHTLLISMVGTITGLPTGLLIGIPTAPAKHKVAALQK 342
 + + +G+ +T+L+ + GL G + I + + K
 Sbjct: 295 QANPNDFLITLFRNLFGSILTVLLTAPSVFFGLIGTGVALLISDI-----K 344

45 Query: 343 LFGWLTIVYIRIFPGTFMIVQSMVIYGTAPF-----GISIDRTLAIPIVGINQNM 397
 + IY+E FRGTFM+VQ +IY+Q F GI+IDR AAI +G+N AY+
 Sbjct: 345 PLQLIFRIYVEFFRGTMLVQLPIYRGLPALFKIQLGTTIDRFAAIIASLNVAAZL 404

50 Query: 398 SEIVRGGIFAVDKGQFKATALGFTHQGTMRKIVLPQVVRNILEPATNGEFVINKDTSVL 457
 +EI+RGI +D+GQ+ +A +LG + QTM+++ PQ R ILP GIEP+ IKITS+
 Sbjct: 405 ASIIKGIGSIDQGNEACEGLMSFMQTMKEVIFPQAFRILPLPGHEPTILIKDTSLT 464

Query: 458 NVISVVELYPSGNTVATQYQYQPTFTLAIIVFLVFTFTVRLIYIERFDF 509
 VI EL+ G + TY+ P+ +A+Y +LT + + + +E D
 Sbjct: 465 AVIGQELFREQLIVATTYRAFEVYIAVALVLLLTISFVFKNLNVMYD 516

An alignment of the GAS and GBS proteins is shown below:

Identities = 82/210 (39%), Positives = 113/210 (53%), Gaps = 12/210 (5%)

55 Query: 14 WGAFPHNFDLPFGFLYVGLISFGALLLALIGLISGLST--SKVKVL-----I 63
 W F ++ F +G TL IS + L +G+L G T+ K KV L + +
 Sbjct: 288 WAIFKGNWQFLRGTHLLISMVGTITGLPGLLIGITPAPAKHKVAALQKLPQWL 347

60 Query: 64 SRIYVEFQNTPLLVQWVFYTGALISGVMISAFFPIALCVGLYHGAIVISVIRSGI 123
 IY+E+P+ TP+VQ + +YFG A + I A+ V + GAY+SK+R GI
 Sbjct: 348 LTIYIEIFRGTFMIVQSMVIYGTAPF--ISIDRTLAIPIVGINQYMSVIRVSGI 405

Query: 124 EAVFPGQTEAALAQGTANQIMQILPQAVRTILPMDGVNVLINVTVAISGADI 183
 AV RQQ +A A GFT QIM+ I+LQ VR ILP N+ V IKTS + +IS ++
 Sbjct: 406 FAVDKGQFKATALGFTHQGTMRKIVLPQVVRNILEPATNGEFVINKDTSVLNVISVVEL 465

65 Query: 184 MFVAKAWAYDTNVIYAPAGAAIYFVVIC 213

-465-

F A T Y F A I YFV+ F
 Sbjct: 466 YFSGNTVAIQYQYFQFTTILALIVFLVF 495

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 367

A DNA sequence (GBSx0398) was identified in *S. agalactiae* <SEQ ID 1195> which encodes the amino acid sequence <SEQ ID 1196>. This protein is predicted to be amino acid ABC transporter, permease protein. Analysis of this protein sequence reveals the following:

Possible site: 39
 >> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -6.95 Transmembrane 25 - 41 (16 - 42)
 INTEGRAL Likelihood = -3.61 Transmembrane 66 - 82 (65 - 86)
 INTEGRAL Likelihood = -2.44 Transmembrane 184 - 200 (182 - 201)
 INTEGRAL Likelihood = -0.59 Transmembrane 119 - 135 (119 - 135)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.3781(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB14704 GB:Z99118 glutamine ABC transporter (integral membrane protein) [Bacillus subtilis]
 Identities = 84/206 (40%), Positives = 129/206 (61%), Gaps = 6/206 (2%)
 Query: 10 ILELLQGFGLTLYISFISILLSMFFGILLAMRNKNPKIWLKIASIYIEFVRNVENLMI 69
 + FL GF +TLX++FISIL+LS PFG + +R +K P+ + + +E +RN+P LL I
 Sbjct: 12 LAPFMDGFLVTLTVAFISILSFPFGLIAGTIRYAKVPVLSQLIALVETIRNLELLII 71
 Query: 70 FIIFLVF-----QMKSVSAGITSFTIPTSAAALAEIIRGGLGVGDKGQTEAGLSQGFYTLQ 124
 F F + + + +A IT+ TIF SA L+EIIR GL +DKGQ EA S G +Y Q
 Sbjct: 72 FTFFFALFEIGIKLEITAAITALTIFESAMLSIEIIRSGLSIDKGQTEAARSSGLSYTQ 131
 Query: 125 VFIIIIIFPOAFKMLPAIISQFVTVIKDTSLLYSVIAIQEIFGKSQILMGRYFEAGQVFT 184
 I+ PQA R+M+P I+SQF+++KDTSL VIA+ E+ +QI+ G+ + F
 Sbjct: 132 TLFFIVMPQALRRMVPPIVQSIFSLKDTSLAV-VIALPELIHNAQIINGQSDGSYFFP 190
 Query: 185 LYAIITAVYFITNFIISPSFKLSKR 210
 + + + +YF N+ +S +R+L R
 Sbjct: 191 IFLLAALMYFAVNYSLSLAARLIEVR 216

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1197> which encodes the amino acid sequence <SEQ ID 1198>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood =-10.51 Transmembrane 529 - 545 (517 - 551)
 INTEGRAL Likelihood =-10.30 Transmembrane 697 - 713 (693 - 719)
 INTEGRAL Likelihood = -4.41 Transmembrane 560 - 576 (555 - 585)
 INTEGRAL Likelihood = -0.32 Transmembrane 662 - 678 (662 - 678)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.5203(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAAL7584 GB:D90907 glutamine-binding periplasmic protein

-466-

[Synecocystis sp.]

Identities = 153/475 (32%), Positives = 251/475 (52%), Gaps = 27/475 (5%)

5 Query: 273 IVDSSSPAFFPQN-GKGKYVDIIEIKAIKQQGPKIEIANPQDAALNAVQSSQADG 331
 +++ +F PFE + G+ G D++LI+AI + ++I ED + A+QS+
 Sbjet: 46 VATEPIFPFFMDEATQGLTGFDVLLIQAIGEAQVTDVQSYFPGDGIIPALQSTVGA 105

10 Query: 332 VIAGATITDARKAIFDFSDPYTTSNIIILAVKAGN-IKNYEDLRKTVGAKNGTSSYSWL 390
 I+ TIT R FS PY+ S +AV+ G + IKN +DL+ K + GT+ + +
 Sbjet: 106 ALSAITITTFERAQSVSPSSPYFKSVLAIAVQDNDTIKNLKDLEGRKLAIAVGITG-AMV 164

15 Query: 391 KENAPKYGIVKAFDDGSSMYDSLANSQSVLAIMDEAVLKTAISQ--RRFPTPLBIST 448
 N P G V FD +S L +G+ DA++D VL YAI R + + S
 Sbjet: 165 RINVP--GAKVINFDISITSAIQELVINGNADAVINDRPVLLIATKDAGLRNVKISADVGSE 222

20 Query: 449 GSVGPAVKGTNPGLI---EMFNGLAALKSSQYDDIIDKYLDSKKA----ATPSBEG 500
 G A+ E+ E+ N GL + +G Y+ I +K+ K PS G
 Sbjet: 223 DYGIAIPLAPPGEINQTRVLAQGLFQIENGTYNAITKRWGKRNPPFLPLVAESLVG 282

25 Query: 501 -----ADESTISGLLSNYKQLIAGLSTTSLTSLISFAIIIGIIPGMVAWSP 549
 + + L + L G T+ LT S +I G + +S
 Sbjet: 283 KVGTAQSLTERQANPNFLITLFPNLFKGSLITVLLTAFPSVFFLGIGTGVAIALISD 342

30 Query: 550 TKSRLISITVFDVVRGIPLMIVAAPIFWGVNPLIESMTGHQSPINDFLAATIALSANG 609
 K L+LI ++V+ RQ P+++ I+G+P L + + G I+ F AA IALGIA
 Sbjet: 343 IKPLQLIFRIYVEFFRGTPMLAQFLITYFGLPALFKZI-GLGITTDIFRAIIIALSLNA 401

35 Query: 610 AVIATIRGGIEAVPAGQWEASRLGSLGYTMRKVLPQAVKMLPNFIPQVLSLAUT 669
 AV+AEI+RGGI+++ GQ EA ELG+S TM++VI PQA + +P N+P+ +DIT
 Sbjet: 402 AYLAETIRGSIQSIDQQWEACSLGMSFWQTMKEVIPPQAPKRIPLPNEPTITLKOT 461

40 Query: 670 TIVSALGLVLPQTGKI IIRNYQSPRMAYLAIYILMIILLTLAKRLEKIN 724
 ++ + IG ELF+ G++I+A Y++F +Y A++YL++ + + K LE ++
 Sbjet: 462 SLTAVIGFCELFREQVLIVATTYRAFEVIAVALVYLLTTISSFVKENLYMD 516

45 Identities = 68/247 (27%), Positives = 106/247 (42%), Gaps = 11/247 (4%)
 Query: 7 VILLAIIMSIFLTCHIASAETIAVSDTAYAPPEKD--SDQIYNGIDVQIINEVAKRQSW 64
 VLL + + + S +TI + + + PFE D + Q+ G DVD+I + +
 Sbjet: 24 VLLAIAIPLLPAPSVSRQTIIVATEPTFPFFMTDEATGQL-TGFDVLLIQAIGEAQV 82

50 Query: 65 DPFMSFGPFDAAVNVAQSQASALMAGTTITMARKKVPHFSEPTTCKIVATRIKAN-AI 123
 + FD + A+QS A ++ TIT R + FS PY+ + + IA + N I
 Sbjet: 83 TVDIQYFPDGIIPALQSNVGAALSAITITFERAQSVSPSSPYFKSVLAIAVDQNDTI 142

55 Query: 124 KKYSLDKGTIVGVKNGTAQAQFNKKKDYITVTFDTGDLMYNLSLPGSIAAVMDORA 183
 K DL+GK + V GT N V FD+ L G+ AV+D
 Sbjet: 143 KNLKDLGSKRLAIAVIGTTGAMVATNVP---GAKVINFDISITSAIQELVINGNADAVINDRP 199

Query: 184 VIQYALS---QNODIANMKGEPIGPSFGFAVKKSGSYDYLVDNFNTALKAMADGTQYR 239
 V+ YAI -N I+ ++ E + + N L + +GT Y A
 Sbjet: 200 VLLYAIKDLGRNVKISADVGSEDDYGIAPPLAPPGEINQTRVLAQGLFQIENGTYNA 259

Query: 240 IMTFKMLG 246
 I KW G
 Sbjet: 260 IYKKNFG 266

An alignment of the GAS and GBS proteins is shown below:

Identities = 68/210 (32%), Positives = 113/210 (53%), Gaps = 16/210 (7%)

60 Query: 13 LQQQGLTLYISYIISLLSMFPGLTALIMNSKNPKMLIASIYIEFVNENVLIMIIPII 72
 LL G G TL ++ IS +++ G + +M S +LI++++++ VR + P + + I
 Sbjet: 517 LLAGLSTTSLTSLISFAIIIGIIPGMVAWSPFKSLRLISTVFDVVRGIPLMIVAAP 576

65 Query: 73 F-----LVPMKMSVAGITSFTFT-----SAALAEIRGLNGVDKQTEAGLSQGV 120
 F L+ M + I F T A +ARI+GG+ V QQ EA S G
 Sbjet: 577 FPGVFNLIEMTGHQSPINDFLAATIALSANGAYIAETIRGSGIEAVPAGQWEASRLG 636

-467-

Query: 121 TYLQVFIHIIFFPQAFKRLPAIIISQFVTVIKDTSLLYSVIAIQSIPGKSQILMGVFEAG 180
 +Y +I PQA +MLP I+QFV +KDT++ S I + E+F +I++ R +
 Sbjct: 637 SYGTTMRKVLIPQAVKLMIPNFINQFVLSKDTTV-SAIGLVELPQTGKIIIRNY--- 692

5 Query: 181 QVFTLYAIITAVYFITNFISSFSRKLKR 210
 Q F +YAI+ +Y I +++ ++L KR
 Sbjct: 693 QSFMYAILAIYILIMILLATLAKRLEKR 722

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 368

A DNA sequence (GBSx0399) was identified in *S. agalactiae* <SEQ ID 1199> which encodes the amino acid sequence <SEQ ID 1200>. Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -12.21 Transmembrane 7 - 23 (1 - 30)

----- Final Results -----
 bacterial membrane --- Certainty=0.5883 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GF:BAB04094 GB:AP001508 unknown conserved protein in B. subtilis
 [Bacillus halodurans]
 Identities = 43/157 (27%), Positives = 83/157 (52%), Gaps = 9/157 (5%)

Query: 26 YQSQPKNTNQAIAIAYKDAKVAKK--DVIHQKIDKEFENFRGSYEIEFNTKSAEYSYHV 83
 +Q++ N+ L +A ++ + + + +K+ +N R YEIE EY + +
 Sbjct: 38 HQASSVSADNGLTLAEASDIALERAGNGVVTEAKDRDNGRVVIEIEVNDDEYDFKI 97

Query: 84 DVKTQQLERDMNNGFSKSTQSQSSSSSQSKSHKIQBEAKKIAFDKNIIESESVNLSKI 143
 D +TG+LL+ + SK SSS ++ IS +BAK+IA K+ + ++ +++++
 Sbjct: 98 DQQTGEILLKKEQQRKSKPREGHSSSKGSEA-VISMDEAKETALKEVS---GKIDDEL 153

Query: 144 KEEIENGKSVYDIDF-VDLQKQEVVDYQIDASTGKII 179
 E ENG VY+++ D + + +V +DA TG ++
 Sbjct: 154 --ERENSLAVIEVEISDHYDDQDVTYVDAMTGNVL 188

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1201> which encodes the amino acid sequence <SEQ ID 1202>. Analysis of this protein sequence reveals the following:

Possible site: 57
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -5.15 Transmembrane 42 - 58 (41 - 60)

----- Final Results -----
 bacterial membrane --- Certainty=0.3060 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below:

Identities = 37/96 (38%), Positives = 63/96 (65%), Gaps = 5/96 (5%)

Query: 94 DMNNGFSKSTQSQSSSSSQSKSHKISQBEAKKIAFDKNIKESVSNLKIKKEIENGSKV 153
 DMD+ +Q +S + K K+S+++AK IA KDS++ E++ L + ++ E+GK+V
 Sbjct: 59 DMDKD-DHMDNQPKTSQTSKVKLSEDKAKS IALKDASVFRADQMLSVTQCQEDGKAV 117

-468-

Query: 154 YDIDFVDLKNKN-EVDYQIDAETGKLIERSDHFMD 188
Y+I+F +NK+ E Y IDA +G I+E+S + +ND
Sbjct: 118 YEIRF---QNKDQEYSYITDANSGDIVEKSSSEPIND 150
Identities = 23/62 (37%). Positives = 37/62 (59%)

Query: 35 NQALAIAIYKDAKVAKKDVIHQKIDKEFENFRGSGYEIEFNTKSAEYSYHVDVKTGQILERD 94
 ++A +IA KDA V + D + ++ E+ + YEIEF K EYSY +D +G I+E+
 Sbjct: 85 DKAKSIALKDASVTEADAQMLSVTQDNEDGKAVYEIRFQNKDCEYSYTTIDANSBGDIVEKS 144

Query: 95 MD 96
+
Subject: 145 SE 146

A related GBS gene <SEQ ID 8563> and protein <SEQ ID 8564> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1      Crend: 9
McG: Discrim Score: 14.45
GVH: Signal Score (-7.5): -5.92
Possible site: 39
>>> Seems to have an uncleavable N-term signal seq
ALON program count: 1 value = -8.92 threshold: 0.0
    INTERNAL Likelihood = -8.92      Transmembrane 7 - 23 ( 2 - 28)
    PERIPHERAL Likelihood = 10.93    37
modified ALON score: 2.28

*** Reasoning Step: 3

```

```

----- Final Results -----
      bacterial membrane --- Certainty=0.4567(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

26.1/59.2% over 140aa
Bacillus subtilis
 EGAD|107494| hypothetical protein Insert characterized
 GP|2632048|emb|CAA05607.1|AJ070271 YkoJ Insert characterized
 GP|263362|emb|CAB13185.1|J99110 similar to hypothetical proteins from *B. subtilis*
 Insert characterized
 FIR|F69859|F69859 conserved hypothetical protein ykoJ - Insert characterized

```

O0905925(379 - 852 of 1164)
EQAD|107494|BS1329(29 - 169 of 170) hypothetical protein [Bacillus subtilis];
GP|2632048|emb|CAA05607.1|]YkoJ (Bacillus subtilis);
GP|2633682|emb|CAB1385.1|]Z99110 similar to hypothetical proteins from B. subtilis
[Scillus subtilis] PIR|P69859|P69859 conserved hypothetical protein ykoJ - Bacillus
subtilis
Search = 6.2
%Identity = 26.1 %Similarity = 59.2
Matches = 37 Mismatches = 52 Conservative Sub.s = 47

```

297 327 357 387 417 447 468 498
NIIE**KEGSCMIKKKVFLEVLVLLVGLGGVLPYQSFQRTINQALAIAYKDAKVAKKDVIH---QKTDKFEFNFRNG
 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
MLKKKVVGVLLAGCLAGGFSFYNAFATSNWENRQRASKTDALTQSQAALAKTIVLDGTVDREDRLDYLNGE

10 20 30 40 50 60 70

[illegible]

65

```
762      792      822      852      882      912      942      972  
ERTKNGSKSVYDIDFVLGLKNKNEVDYQIDAETGCIERSRDHMDN*FK*DIKKRRSKRPFS*LGLSSLLFT*KPT*KT*DD  
| : | : | : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
ED--DGAYIYEMR-IQTCKGTETEFSISAKDGRILKEIQIDD
```

SEQ ID 8564 (GBS37) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 14 (lane 4; MW 22kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 10; MW 47kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 369

A DNA sequence (GBSx0400) was identified in *S. agalactiae* <SEQ ID 1203> which encodes the amino acid sequence <SEQ ID 1204>. Analysis of this protein sequence reveals the following:

```

Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1499(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9709> which encodes amino acid sequence <SEQ ID 9710> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1205> which encodes the amino acid sequence <SEQ ID 1206>. Analysis of this protein sequence reveals the following:

```

Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2808(Affirmative) < succ>
bacterial membranes --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

Identities = 128/297 (43%), Positives = 180/297 (60%), Gaps = 9/297 (3%)

Query: 54	IIDDKVIVKIPKYPWT-GLSLQAPLALFLEVLQNRMPISLSEGLSEFTLQVLPAPDRLLA	112
	I D + + GSP F W Q + L F P + L + M P L E L G + E T Q L A + P	
Sbjct: 46	IIDNLGSSPTFWVWPIEKNGSAKQLLTLLDLVEMPELSSQLHETQLTLTTPPSLL	105
Query: 113	PHDDFWIALSQVIYDSFPGNSIALSDTVLNKLIHQFRLISQQAQVVRYPKDVGTGDR	172
	P F W L + + + FVGG + L + L + + L H Q F R + I S S Q A G + + + K + M D	
Sbjct: 106	PCHMFWEKLSAIVDPAGFGKLSQAGLEKRLAQKRFVLSQQAQGISRNMYIKMELTQ	165
Query: 173	ALVNL + + + -	226
	AL + L + K K C K P L Q R + + S A R L I H N K R + P P E N K I L F I T E	
Sbjct: 164	ALALFSLRKSGKPCMLQAPDYTLMDGAELENLKLRDENVKLVPSQVSNIKVLVFI	225
Query: 227	FIIDDRGFNLSDAEVITRNGLINGASPNYAQTRHKELDVDVPC-LDKPRFMDTR	285
	F + D G P L N E - D A E V T - G I N G A S P N Y + R H + L D V D P H - D E P R D +	
Sbjct: 226	F + I D S T G F F L N E L D A E V T E K I G V A S P N Y - T G D R F M D V D P I S H D G P R D T L	284

Query: 266 GYRSPNLSRRKWNFFKRDYDCSYFNKKGYAAGRSSAKQSDVKYKYLKAVQOMR 342
 G+RSP R+WF ++D+ SYFN RG +A+ +S+ V K K K+ + ++
 Sbjct: 265 GFRSPKRVGFQVFRQKDDPMFSYFMAGLFAYHNKSSFARVQKSAQKFKRQIPIK 341

5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 370

A DNA sequence (GBSx0401) was identified in *S. agalactiae* <SEQ ID 1207> which encodes the amino acid sequence <SEQ ID 1208>. This protein is predicted to be similar to two-component response regulator [YcbM] (ompr-like protein). Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3129 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA55264 GB:X78502 gtcR [Brevibacillus brevis]
 Identities = 99/228 (43%), Positives = 149/228 (64%), Gaps = 3/228 (1%)
 Query: 2 RTVLVVGQDDTIELLRSTYLGALYKVVWASDGRHAFSLPQHQIDLAIDITLPKIDGY 61
 +T+L+ + E IELL+ +LE Y+++ A DGE+AA+ +QH +DLAIDDI +P +DG+
 Sbjct: 3 KTLIADDEPEITIELLKLFLERESYRIIRAYDGBQANNYIRQHPVOLAIDIMFALDGF 62
 Query: 62 ELTRLIRQDSQIPIMLAAKTIDMRILGILNIGADDFTKPFNSLEVLIARINSQIRYYE 121
 +L + + + ++P+I+LAK D D+ILGL +GADDFI+KPFN LE +ARI +QLAR +E
 Sbjct: 63 QLIKRLITNRYKLPVILISAKNRSDKILGLGLGADDFISKPFNPLEAVARIQARAPAE 122
 Query: 122 FNSLAKP---KNQFIKIGELDEEHVELTNGKHILKTATFEKILHILMS-SPGRIYTKT 178
 FN + Q +G L L + + + +T E++IL+ M S I+TK
 Sbjct: 123 FNPKEKAISTQSTIVGRLLTLLTACVYRGDETYSVTPLEYRLNTFMQCSRTSPITQ 182
 Query: 179 QLVKIKINGRYLEGDETTIMVHISNIRDKIEDSKYPKYIKTLRGVGYK 226
 QL+E+ D+ TIMV IS +RDKIED + P YIKT+RG+GYK
 Sbjct: 183 QLFQAMSETYWEDNTIMVQISRLDKIEDQPSQPVYIKTVRGLGYK 230

There is also homology to SEQ ID 1182:

Identities = 87/230 (37%), Positives = 144/230 (61%), Gaps = 5/230 (2%)
 Query: 1 MRTLIVVGQDDTIELLRSTYLGALYKVVWASDGRHAFSLPQHQIDLAIDITLPKIDGY 60
 M+ +L+V + +++++ L Y +V A DG EA +F++ + DL I+D+ LP+DG
 Sbjct: 1 MKKILIVDDEKPISDIITKPNLTKSGYDIVTAFDGRNAVITFEZEKPDILIDLMPELDG 60
 Query: 61 YELTRLIRQDSQIPIMLAAKTIDMRILGILNIGADDFTKPFNSLEVLIARINSQIRYYE 120
 E+ + IR+ S +PIIIL+AK ++ D++GL IGADD+TKPFE+ E+LAR+ + LRR
 Sbjct: 61 LEVAKETIKTSHVPIIMLSAKDSEFFKIVGLEIGADDYVTKPFSNRELIARVKAHRETE 120
 Query: 121 FNSLAKPKN-----QFIKIGELDEEHVELTNGKHILKTATFEKILHILMS-SPGRIY 175
 + +N Q + IG L++ + K+G+ ++LT EP++LH L + G++
 Sbjct: 121 TITFAVEENASSGTQLTIGNLQILPDAPVAKKHGQREVELTHREPELLHLLHNSQGVW 180
 Query: 176 TKPOLYKINGRYLEGDETTIMVHISNIRDKIEDSKYPKYIKTLRGVGY 225
 T+ L E + G GD T+ V + +R+KIED P+YI T RGVGY
 Sbjct: 181 TREHLLIETVWGYDYFGDVRTVDVTVRRLREKIDTTPSRPYIILTRGVGY 230

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 371

A DNA sequence (GBSx0402) was identified in *S. agalactiae* <SEQ ID 1209> which encodes the amino acid sequence <SEQ ID 1210>. This protein is predicted to be threonyl-tRNA synthetase I (thrS). Analysis of this protein sequence reveals the following:

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.2353 (Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB06860 GB:AP001517 threonyl-tRNA synthetase 1 [Bacillus halodurans]
Identities = 413/638 (64%), Positives = 506/638 (78%), Gaps = 7/638 (1%)

Query: 1 MIKITFPDGAIREFESGITTFEIACSSISNLAKKALAGKFNGLDITTRAIEDGSIIEV 60
MI ITTFDGA++EF G TT EIA SIS L KALAG +G L+D IE+DG+I IV
Sbjct: 4 MINITFPDGAIVKEFGKTTTABIAGSSISPLKKKALAGMLDITLIDMTIEQDGTITIV 63

Query: 61 TPDHEDALGVLRHSAHLFAQAARLFPD--LCLGVGPALQDGFYYDTDNKSGQISNDL 118
TP+ ++AL VLRHS AH+ AQA KRLF D + LGVGP I+ GFYYD D ++ +DL
Sbjct: 64 TPSEDEALEVLRHSTAHVMAQALKRLFQDRNVKLGVGVPVIEGGFYICVDNDES -ITPBDL 122

Query: 119 PRIIEEKKIKIKENHPICIRIEESKEEALSLFYD--DPYKVELISEHADG-LTVYRQGEF 175
P+IE+EMKCI+ EN P R +S+EEL +++ DPYK+SLI++ ED +T+Y QGEF
Sbjct: 123 PKIKEMKKIIGENLPIERVVVSREELARYEVDGDPYKIELINDLFEDETITTYEQGEF 182

Query: 176 VDLCRGPHVFTSGRIQVTHLANVAGAYWRGNSDNAMQVRVGTAWFDKKLKYLRREE 235
DLRG HVPSTG+++ F LLN+AGAYWRG+S N M+QR+YGT+ F K DL +L+ EE
Sbjct: 183 FDLCRGPHVFTSGRLKRLKLANLAGAYWRGSSNMLQRIYGTAFKKADLDEHLELLE 242

Query: 236 AKERDHRKLGKELDLFWNVPEVGGQLPFWLPGATIRRELRYIVDKELASGYQHVTYP 295
AKERDHRKLGKEL +F ++ +VGGQLP WLP GATIR +ERYVDKE GYQHVTYP
Sbjct: 243 AKERDHRKLGKELGIFALSQKVGQGLPFWLPGGATIRRIERYIVDKELASGYQHVTYP 302

Query: 296 NASVRFYKTSGEWDHYREDMFPTMDMGDGEFVLPRMNCPHIEVYKHVHSYRELPIRI 355
+AS E YKTSGEWDHY++DMFPTM+M + EE VLPRMNCPHI+ VYK +SYR+LP+RI
Sbjct: 303 LASSLEFKTSGEWDHYKDDMFPTMEN-ENELVLPRMNCPHIMVYKTYETMSRYQLPRLI 361

Query: 356 AELGMMHRYEKSALTGLQVRVEMTINDAHIPTVPEQIKDEFKALANLAIIEYEDFNLD 415
AELG+MERYE SGA++GLQVRV MTINDAHIF +QIKDEF++ +LI +YEDF L +
Sbjct: 362 AELGMMHRYEKSALTGLQVRVEMTINDAHIPTVPEQIKDEFKALANLAIIEYEDFNLD 421

Query: 416 YRFLSYRDPEDKHKYINDENMENAQAMLEKAMDDGLDYFEARGAAFYGPGLDIQVK 475
Y RFLSYRDPEDK KY+D+D MW AQ MLKEAMD+ L+YFARGAAFYGPGLD+QV+
Sbjct: 422 YSFLSYRDPEDKHKYINDENMNAQAMLEKAMDELRLKYFEARGAAFYGPGLDQVK 481

Query: 476 TALGNESTLSTIQLOFLFLPKRFDLKYIGADGEGHRPIMIHGGISGMERFTAILISTYK 535
TALG +ETLST+GLDFLFLPKRFDL YG D+G HRP+++HGG +STMERF A L+E YK
Sbjct: 482 TALGKDETLSTVQLFLFLPKRFDLTVYSGDQGHRRPVVHVGVSVMKEKFAFLLETKG 541

Query: 536 APTTWLAPQQVVISPISEAHIDYANFVARVLKDRGIRAEVDNERMRYKIRAAQTQIKI 595
APTTWLAP QV VIP+S EAH++YA V L+ GIR E+D+R+EK+ YKIR AQ QKI
Sbjct: 542 APTTWLAPVQVQVIVPSPEARLEYAKNVQETLQAGIRVIEDERDKIGYKIRAAQMQKI 601

Query: 596 PYQLVIGDKEMEKAVNVRRYSGKATETKSIIEFVESI 633
PY L++GDKE+E VNVK+YG K + + +EFV +
Sbjct: 602 PMLVLIGDKVEANGVNVRRYSGKSSMLDEFEVRIV 639

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1211> which encodes the amino acid sequence <SEQ ID 1212>. Analysis of this protein sequence reveals the following:

-472-

Possible site: 32
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 5 bacterial cytoplasm --- Certainty=0.2566(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

- 10 Identities = 566/644 (87%), Positives = 608/644 (93%)
- Query: 1 MIKITFDGAIREFESGITTFFIAQGISNSLAKKALAGKFNQGLIDTTRAIIBEDGSIIEV 60
 1 MIKITFDGGA+REFESG+TF+IA+SIS SLAKKALAGKFN QLIDTTRAIIBEDGSIIEV
 15 MIKITFDGAVREFESGVTTFDIARISKSLSAKKALAGKFNQGLIDTTRAIIBEDGSIIEV 60
- Query: 61 TPDHEDALGVLRHSAHLFAQAARLFFDLCLGVGPAIDGFIYDTDNKSGQISNDLDR 120
 61 TPDH+DA VLRHSAHLFAQAARLFF+L LGVGPAI +GFYYDTIN QGISN+DLDR
 15 TPDHKDAYEVLHSAHLFAQAARLFFNLHLGVGPAIABGFYYDTDNABGQISNEDLDR 120
- Query: 121 IBSEMKKIVKENGHPCIRBEISKKEALELFKDDFYKVELISEHAEDGLTVYRQGFVDLCR 180
 121 IB EM+KIV EN+PCIRE++KKEALELFKDDFYKVELI+EHA GLTVYRQGFVDLCR
 15 IBEMKKIVIRMYPCIRBEVTKKEALELFKDDFYKVELINEHAGAGLTVYRQGFVDLCR 180
- Query: 181 GPHVPTSGRIQVPHLLNVAGATYMRGNSDNNMQRVYGTAMFDKKDLKAYLRKEEAKERD 240
 181 GPHVPTSGRIQVPHLLNVAGATYMRGNSDN MQR+YGTAMFDKKDLKAYLR EAKERD
 25 GPHVPTSGRIQVPHLLNVAGATYMRGNSDNNMQR IYGTAMFDKKDLKAYLTR EAKERD 240
- Query: 241 HRKLGKELDLFMWNPVVGQGLFPWLPNGATIRRELSRYTDKEIASGQYQVITPPMASVE 300
 241 HRKLGKELDLFM++ EVOQGLFPWLP+GATIRR LERYT DKE+ASGQYQVITPP+ASVE
 30 HRKLGKELDLFMISQGEVQGLFPWLPDGTATIRKTLERYITDKEIASGQYQVITPPASVE 300
- Query: 301 FYKTSQGHWDHYEDMFFTMKGDGEFVLRPMNCPHHIEVYKHHVHSYRELDIRIAELGW 360
 301 FYKTSQGHWDHY+EDMFF MKMGDGEFVLRPMNCPHH+VYK+HV SYRELDIRIAELGW
 35 FYKTSQGHWDHYQEDMFFVMKMGDGEFVLRPMNCPHHIVQYKHHVRSYRELDIRIAELGW 360
- Query: 361 MHRYEKSGALGLQVRVEMTLMDGHIFVTPPEQIKDEFKALANLIARTIVEDFNLIDYRFL 420
 361 MHRYEKSGAL+GLQVRVEMTLMD HIPVTPPEQI++EF +AL LI ++Y DFMIDYRFL
 40 MHRYEKSGALSGLQVRVEMTLMDGHIFVTPPEQIKDEFQRAQLIIDVADFMIDYRFL 420
- Query: 421 SYRDPEDKHKKYDNDDEMENQAQMLKRAMDDGGLDYFZABGEAAFYGPKLIDIQVKTALGN 480
 421 SYRDP D HKYDNDDEMENQA+MLK A+D+ G-DYFZABGEAAFYGPKLIDIQVKTALGN
 45 SYRDPDCTHKYDNDDEMENQAQMLKALIDEMGVDFZABGEAAFYGPKLIDIQVKTALGN 480
- Query: 481 ESTLSTIQLDPLPERFDLAKYIADGGEHRPIMIHROGISTMERFTALLIETKGAFFTW 540
 481 ESTLSTIQLDPLPERFDLAKYIADGGEHRP+MIHRG ISTMERFTALLIETKGAFFTW
 50 ESTLSTIQLDPLPERFDLAKYIADGGEHRPVMIHROGISTMERFTALLIETKGAFFTW 540
- Query: 541 LAPQGVSVIPISNEAHIDYAMEAVARVLKDRGINARVDERNEKMQYKIRAACTQKIPVQLI 600
 541 LAP QV+VIPISNEAHIDYAMEVA+ L+DRG+RA+VDERNEKMQYKIRA+CT KIPVQLI
 55 LAPHQVTVIPISNEAHIDYAMEVAKTLDRGVADVDNERNEKMQYKIRASCTKIPVQLI 600
- Query: 601 VGDKEMEKAVNVRRYSGKATETKTSIEFVESILADIARKSRPD 644
 601 VGDKEME+K+NVNRRYSGK T T+S+REFVE+ILADIARKSRPD
 55 VGDKEMEDSVNVRRYSGKATETKTSIEFVESILADIARKSRPD 644

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 372

- 60 A DNA sequence (GBSx0403) was identified in *S.agalactiae* <SEQ ID 1213> which encodes the amino acid sequence <SEQ ID 1214>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1985 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA72250 GB:Y11463 ORF5 (Streptococcus pneumoniae)
 Identities = 189/290 (65%), Positives = 234/290 (80%)

Query: 1 MRIGLFTDTYFPQVSGVSTSIKTLKGLKRGHEVYIFTTDRNVRKFPDPTIIRLPSVP 60
 MRIGLFTDTYFPQVSGV+TSIRTLK LEK+GH V+IFTTID+V R+ED IIR+PSVP
 Sbjct: 1 MRIGLFTDTYFPQVSGVATSIKTLKTELEKQGHAVFIFTTIDKDVNRYEDWQIIRIPSPV 60

Query: 61 FISFTDRRVVYRGILISAYRIAKDYELDIHTQTEPSGLGLKLVAKALRIPVHTYHTQY 120
 F+F DER YRG A IAK Y+LDIHTQTEPSGLGLS +A+ L+IPV+HTYHTQY
 Sbjct: 61 FFAFKDRRFAYRGFSKALEIAKQYQLDIHTQTEPSGLGLGIWIAELKIPVHTYHTQY 120

Query: 121 EDYVGYIANGKLIKPSMWKYIMRTYLSDLGVCPSRIVMLLDGYGVKIPKQVITGIP 180
 EDYV YIANG LI+PSMWK+R +L D+DGVICPS IV +LL Y VK+ K+VITGIP
 Sbjct: 121 EDYVHYIANGMLIRPSMWKYLVRGLHVDGVICPSRIVRDLSDYKVKVEKRVITGITE 180

Query: 181 VENVYRRDISSETIKNLKTGLADNMLSLSRVSFEKNIQAILMHLASVDNPHVK 240
 + + R +I +E +K LR++LG+ D + LLSLR+S+EKNIQAL+ +V+ E VK
 Sbjct: 181 LAKFERPEIKQNLKELRSLKLSIQDCKTLKLSRISYENKIQAVLWAFADVLEEDKVK 240

Query: 241 LVIVGSGPYLQDLKGLVHSLSELENSVIFTGMVHESQVAILYKACDFFISA 290
 LV+ GDGPYL+DLKE +LE++SVITGM+ S+ A+YKA DFFISA
 Sbjct: 241 LVVAGSGPYLADLEQAQNLKIQDGVITGMIAFSETALYKANDFFISA 290

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1215> which encodes the amino acid sequence <SEQ ID 1216>. Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1074 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 309/444 (69%), Positives = 370/444 (82%)

Query: 1 MRIGLFTDTYFPQVSGVSTSIKTLKGLKRGHEVYIFTTDRNVRKFPDPTIIRLPSVP 60
 MRIGLFTDTYFPQVSGV+TSIRTLK LEKRGHEVYIFTTDR+VNRKFPDPTIIRLPSVP
 Sbjct: 1 MRIGLFTDTYFPQVSGVATSIKTLKLEKRGHEVYIFTTDRVNRKFPDPTIIRLPSVP 60

Query: 61 FISFTDRRVVYRGILISAYRIAKDYELDIHTQTEPSGLGLKLVAKALRIPVHTYHTQY 120
 F+SFTDRRVVYRGILIS+Y+IAK Y LDIHTQTEPSGLGLK+ KALRIPVHTYHTQY
 Sbjct: 61 FVSFTDRRVVYRGILISSYKIAKHNLIDIHQTEPSGLGLGKMGKALRIPVHTYHTQY 120

Query: 121 EDYVGYIANGKLIKPSMWKYIMRTYLSDLGVCPSRIVMLLDGYGVKIPKQVITGIP 180
 EDYV YIA GK+I+PSMWK +R YL DLDGVICPSRIVMLL+GY V IPK+VITGIP
 Sbjct: 121 EDYVSYIANGKIIRPSMWKPLIRGYLKLDGVICPSRIVMLLGGYEVITPKRVITGIP 180

Query: 181 VENVYRRDISSETIKNLKTGLADNMLSLSRVSFEKNIQAILMHLASVDNPHVK 240
 +E Y R+DI+ E + NL+ ELG+A +TMLSLR+S+EKNIQAI+ +A+ EN +K
 Sbjct: 181 LKRYIRDDITAEVINLAKELGAGKTNMLSLSRISYEKNIQAILNMPALLAEPAKIK 240

Query: 241 LVIVGSGPYLQDLKGLVHSLSELENSVIFTGMVHESQVAILYKACDFFISA 300
 L+IVG+GPYL DLK L LE++ V FTGM H +VA+YKACDFFISASTSETGITY 300
 Sbjct: 241 LIIVNGPYLQDLKGLVHSLSELENSVIFTGMVHESQVAILYKACDFFISASTSETGITY 300

Query: 301 IESLASGRPIAQSNPYLDDVISDKMPTLYKKESDLADALDAIARPTKMQYAEQKL 360

TESLA93 PIIA NEYLDVV++DKMFGTLY E+L DAI+DAI +TP M + +K
 Sbjct: 301 TESLAGSTPIIAHGNFYLDVVTDKMFGTLYAETDITDAIIDAIAIKTVPMDKRLLAKR 360

Query: 361 YEISAEINFSKSVYAFYLDPLISQKASVKEKVSLTIGNKDSHSTRFVRVGAIVLPDKVFTF 420
 YEISA++F KS+Y FYLD LI++ + +K+SL + + S+L+ V+ A++LPK+
 Sbjct: 361 YEISAQHFKGKSYTFYLDLTIAHNSKEAQLSLSYLHNSGKSSSLKLVQGAHLPKRAKV 420

Query: 421 TGRASKVUVKAPKRRISSIDFELD 444
 T S KVVKAP + + +I+DFELD
 Sbjct: 421 TAITSVKVVKAPIKLVIAIKDFELD 444

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 373

A DNA sequence (GBSx0404) was identified in *S. galactiae* <SEQ ID 1217> which encodes the amino acid sequence <SEQ ID 1218>. This protein is predicted to be lipopolysaccharide biosynthesis protein-related protein. Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4076 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AA919110 GB:AE005009 Vng0600c [Halobacterium sp. NRC-1]
 Identities = 117/350 (33%), Positives = 178/350 (50%), Gaps = 29/350 (8%)

Query: 1 MKVLLYLEAEYLYKSGIGRAIKHQEKALQIAGIDYITNFT----- 41
 M+ L YLEA E L+ G+ A Q AL+ ++ P
 Sbjct: 2 MRALNYLEAAEALR-GGMVATNQRAALETTVVEVVEFWRAGDPVRSGLAAGGSCF 60

Query: 42 DCFDLVHANTYGIKRWLIASKAKKTGKVKIMHGHSTEDFRNFSIGSNLVSPLFKNYLCR 101
 FD+ H N G S + A++T +++H H T EDF SF GS+ ++P + YL
 Sbjct: 61 TAFDVAHCNLYVGGPSVAVARHARRDTPLVLHAHLTFEDFAQSFRGSSIIAPALEPYLRW 120

Query: 102 FYQKADAIITPTDYKQLIKAYGIKKPIFVLSNGIDLSRYORSEKKESAFRHYPHLSKDD 161
 FY +AD ++ P++Y+K +++NY + PI LSNH+DL Q E + R F L D
 Sbjct: 121 FYSQADINLCPSEYTKDVLKAYPVDAPIRQLSNGVLESMQGYESFRADTRARFDL--DG 178

Query: 162 KVVMGAGLYFMKRGIDQFVEVAAKMDIRFIWFGTEINKWVLPRIKRVQIVTKQHPNSVTF 221
 VV G F RKG+ F E+ AK D F WFG ++ + P+NVTF
 Sbjct: 179 TVVYAVGVFPERKGLTMFCLE-AKATDHEFANFGPYDEGQAGAAATRIOWADPPANVTFT 237

Query: 222 GYIKGDVYEGAMSADAFPPSRRETEGIVLEALASHQHVVLRIPIVHWGWT-DSVE 280
 GY++ A A D + FP++ E +GI VLEA+ A + VVLRIPIV+ T+ +
 Sbjct: 238 GYMEDK--RAAFGAGDIYLPFAKVENQGIADVLEMAQCKIVVLRIPIVFRFPFTDGEDCL 295

Query: 281 LATVDVGFVKELDKVLSGSKDKIKRGYH---VAESRSIERIAHSLASVQ 327
 + + + F + +D++ + + G + AFS S++RI ELAS+Y+
 Sbjct: 296 MCSTFATFADRIDLADDPLEKTRKIGENARETASHSIDRIIGELASIVE 345

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1219> which encodes the amino acid sequence <SEQ ID 1220>. Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

-475-

bacterial cytoplasm --- Certainty=0.4088(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below:

Identities = 236/332 (71%), Positives = 276/332 (83%)

Query: 1 MKVLLYLEAEFYLKSGIGRAIKHQKALQIAGIDYITNPDDFDLVHMVTYIGRISWLM 60
 10 Sbjct: 1 MKVLLYLEAE YL+KSGIGRAIKHQ KAL + G +TNP + +DLVH+NTYG+SWLM 60
 Query: 61 SKAKTKGKKVIMHGHSTEDFRNSFTIGSNLVSPLFKWYLCRFYQKADAIITPTIYSKQLI 120
 Sbjct: 61 KA+K GKVKIMHGHSTEDFRNSFTI SNL+SP FK YLC FY KADRIITPT YSK LI 120
 15 Query: 121 KAYGIKKPIFVLSNGIDLSTYQSRSEKESAFRHYFHLSDKKDVVMGAGLYPMRIGDQFV 180
 ++YG+K PIF +SNGIDL +Y KKE+AFR YF + + +KVMGAGL+F+RGID FV 180
 Sbjct: 121 ESTGYKSPILPAVSNGLDLSYQADPFKEAARFYFDI KBSEKVMGAGLFLFRIGDQFV 180
 20 Query: 181 EVAAMPDRIFFIWGETNKGWIPFKVRQIVTKQHPSNVTFAGYIKGDVYEGAMASDAFF 240
 +VA MPD+RFIWGETNKGWIP +VRQ+V HF N+ F GYIKGDVYEGAM+ +DAFF 240
 Sbjct: 181 KVAQAMPDRIFFIWGETNKGWIPAQVRQMVGNGHFKPLI PFGYIKGDVYEGAMIGDAFF 240
 25 Query: 241 FPSRETEGIVVLEALASHQHVLRDIPVYHGWVTEDSVELATDVGDVFKLQKVLGSKS 300
 FPSRETEGIVVLEALAS QH+VLRDIPVY+GW + S ELATD+ GF+E L KV SG S 300
 Sbjct: 241 FPSRETEGIVVLEALASQHVLRDIPVYHGWVTEDSVELATDIFGFIALKKVFSGAS 300
 Query: 301 DKIKGGYHVAESRSTIERIANELASVYQKVMEL 332
 ++K+ GY VA+SR +E + H L YI+KVMEL 332
 30 Sbjct: 301 NKVEAGHKVAQSRLETYGHALVDYVKVMEL 332

Based on this analysis, it was predicted that these proteins could be useful antigens for vaccines or diagnostics.

Example 374

35 A DNA sequence (GBSx0405) was identified in *S.agalactiae* <SEQ ID 1221> which encodes the amino acid sequence <SEQ ID 1222>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5487(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database:

>GP:ARC35010 GB:AF055997 intracellular a-amylase [Streptococcus mutans]
 Identities = 308/483 (63%), Positives = 378/483 (77%)

50 Query: 1 MYNELDQAFWYLPDSGNHKKLETSIDLKKLGISKIWLPPAFKGTSSDDVGYGVYDL 60
 MYNE +MQ FEWYLP+DG HW+ L E S LK +GLEK+M+PPAFKGT S+DWYGVYDL 60
 Sbjct: 1 MYNELDQAFWYLPDSGNHKKLETSIDLKKLGISKIWLPPAFKGTSSDDVGYGVYDL 60
 Query: 61 FDLGEFDQNGTITPTKYGRKSEYLLKLSKANGIKPPADIVLHKGANDHKEKPOVIVN 120
 +DLGEF+QNGT+RTKYG +E+YL + +LK I P +DIVLHKGAND KE+POV+KVN 120
 55 Sbjct: 61 YDLGEFQNGTIVRTKYGSRDYLVNALKGEQDIMPDISDIVLHKGANDAKERPOVIVN 120
 Query: 121 FNNRQSEALSEPYEIRGWTCPPDFGRQGEYNDPKWHWYHFTGLDYDAKNNETDIPMIVEN 180
 F NNQE +SEPYEIRGW T P+PPGRQ Y+DFKNHWYHFTG+DYDA +NR I+M+GWN 180
 Sbjct: 121 FSNRQSEKISEPYEIRGWTCPPDFGRQDQYNSDPKNHWYHFTGLDYDALNENRIMILGN 180
 60 Query: 161 KGWADCDLIDENGNFDYLYMNDIDFKHPFVIVNLDQWAKMFIPTGTIGFPLDAVHHD 240

-476-

KGN A + ID ENGR+DYLMY+DIDFKHPEV ++L+DW WF+ET+G+ GPULDA+KHID
 Sbjct: 181 KGNASQENIDQENGNIDYLMYDIDFKHPEVQEHLDVWVWFLETSGVGGFKLDAIKHID 240
 5 Query: 241 SYFIQTPTINDIRTKIKPDLVPCRYNKSQDQTSMSKDYLEATQPFQSLVQVTLBNPFDASH 300
 F+ FI IR +K DL VPGRYNK + DYI + QF L+DW LHM+ P+A
 Sbjct: 241 KTFGAQPIRYIREHKLADLYVPGRYNKDSHPDITDYIHSDVLDQFDGLDVMLEMSLPEAQ 300
 10 Query: 301 QNRDPMTPTTDDSLVIDNPHYAVTTPVENHDTQSQALSRVDEWFKPLAYGLILLRQGS 360
 + DSD+ TI DDSH+ +P+AVTTPV-NEID+Q QALRS V +WEKPLAYGLILLRQGS
 Sbjct: 301 KGSDFDLSTILLDDSLKSHDFPVAITVDNHDSSQSQALSTVAWEKPLAYGLILLRQGS 360
 15 Query: 361 TPCLPYQDYVYIGSGFQSPKSVKVIDNMRQNVFVQGVQDYPTHSNCIGWTCLODEH 420
 PC+PYQDYVYIG SF Q SP+ V+DK+ +RQ +V+G + T NCIGWTCLODEH
 Sbjct: 361 IPCVPYQDYVYIGSGFQSPKSVKVIDNMRQNVFVQGVQDYPTHSNCIGWTCLODEH 420
 20 Query: 421 NSCLAVLTNDDQGWKHMVEGIEYAGTTPVDYLGNCQGVVIGDDGWDFVIVESASISAN 480
 +AV+ ++NO+ K M +GE K FVDYL NC +EV+ + D GWDGF V+ AS+SAW
 Sbjct: 421 PGGVAVTISNGEANKRMKMGEPNKNVFDYLANCTESVILDQGWDFVQESASISAN 480
 25 Query: 481 VPK 483
 V K
 Sbjct: 481 VNK 483

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1223> which encodes the amino acid sequence <SEQ ID 1224>. Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have a cleavable N-term signal seq.
 30 ----- Final Results -----
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
 35 The protein has homology with the following sequences in the databases:
 >GP:AB00845 GB:M57692 alpha-cyclodextrin glycosyltransferase
 [Thermoanaerobacterium thermosulfurigenes]
 Identities = 356/710 (50%), Positives = 468/710 (65%), Gaps = 16/710 (2%)
 40 Query: 7 KTYKALTKSAVLLGLISPLIT--VSAADNASVINKADPSTDTIYQIVTDRFMDGNTSRNG 64
 KT+KL+ + L L+ F LT + AA + V+N +STD IYQIVTDRF DGNSTNN
 Sbjct: 3 KTFKLLLVMLSLTLV-FGLTAPIQAASDTAVENVVNYSTDTIYQIVTDRFVDGNTSNRP 61
 45 Query: 65 KIDVFDKN--DLKKYHGGDWQGLIATKIDGYLTMGISAIISSPVENIDSIDPN---G 119
 D+D LKKY GGDWQGLI KI DGYLT MG+AIWIS PVENI ++ P + G
 Sbjct: 62 TGDLYDPIHTSLYFGSGDWQGLIINKINDGYLKGWGTAINWISQPVENIYAVLDSFTGG 121
 50 Query: 120 SAAYHGYWAKDFKTIHQEPGTADPQQLKVAHQHHLKVIDEAPNHTSTASKBGTITKE 179
 S +YHGYW+DF +TN +FG+ DFQ L+ AH H+KV+IDEAPNHTS A + T+ E
 Sbjct: 122 STSYHGYWARDPKITNPYFGSFTQFNLTAAHANNKVIIIDEAPNHTSPASSTDTPTAR 181
 55 Query: 180 DGALYKNGKLGWGFSDDKDKI PNHRSWIDSTYENSIYHSMYGLADIANINPKVDQMYR 239
 +G LY NG L+G ++D + F+H TDP+YE+ IY +++ LADLN N +D Y+K
 Sbjct: 182 NGRILYNGTLLGTYTNDTNGYFHYGFTDPSSYEDYGRNIFDLADLNQNSTDIFYSK 241
 60 Query: 240 AIDKWLGLDGVGIRVDVAKHMSQGWKQKNGLSHIYKINNVFPGWSEHSDDDYDMITFA 299
 AI WLD+G+DGIR+DAVKIM GWQK++ I VF PGWGF G + D + T P A
 Sbjct: 242 AIKVWLMGIDGIRLDVAKHMPFGWQKNGFMDSLSYRPFVTFGWGFLGNTSDVNTYFA 301
 65 Query: 300 NNSGMGLIDFRFANALRQLTYGFTPTMRDDFYKLENRDQVINEVTQVTFIDNHDMERF 359
 N GGM LLDPRF+ +RQ++ +T TM ++++ N +D VTFIDNHDM+RF
 Sbjct: 302 NESGMSLLDFRFSQKVRQVFRD-NITDMYGLDSMTQSTASDYNFINDMVTFIDNHDMERF 360
 Query: 360 ATKVANNQYAVNQAYALLLSRQGVNITYGTQYATQDKDPNNRSGMPSFKESQAYIKVI 419
 + V QA A LTRSGVP IYGTQY TG+ DP NR M SFN + AY VI

-477-

	Sbjct: 361	YN--QGSTRPVRGALAFITLTSRGVPAIYYGTBQYMTGNDPYNRAWMTSFTNTSDAXVI	418	
	Query: 420	SKLAPLRKQNALAYGTTSGRWISDHVAVFRRKGNHVALVAINRDQTNQYTTITAKTAL	479	
5		KLAPLRK N A+AYGTT+QRWI++ V ++SRKGN+VALVAINR+ Y IT TAL		
	Sbjct: 419	KKLAPLRKSNPAIAYGTTQQRWINDVYIYERKGNHVALVAINRMLSTSYNTIGLTYAL	478	
	Query: 480	PQNSYKDKLGLGQQLIWAGDGTISSPFGAGQVAVWVYEGEDKTPQLGDVDAVGIA	539	
		F +Y D L GLL G + V +DG+++ F L AG+VAVW Y ++P +G V ++ A		
10		Sbjct: 479	PAGTYTDVLGGLLNGNS:SVASDGSVTPFTLSAGEVAVWVQYVSSNSFLIGHGVGPTWKA	538
	Query: 540	GNKITISGQGFNGSKQVTFGEISARILSWSDTLITLKVTPVANYNISVTTADKQTSN	599	
		G ITI G+GFG +QQV FG + I+SW DZ ++KVE+V YNIS+ T+ TSN		
	Sbjct: 539	QQTITIDGRGFTTSQQVLPSTAGTIVSNDQTFVKVPSVTPGKINISLKTSSGRTSN	598	
15		Query: 600	SYQAFVLTDKQIPVRLINDFKTVPGEQLYMLKDVPEMGANDAKHNVGFLNNQTIAK	659
		+Y ++LT QI VR ++N+ TV GE +XL G+V E+G D A+G+FN Q++		
	Sbjct: 599	TYNNINILTGWQICVRFVFNMASTVYGEVYLTGNVALGNWDTSKAIGMPFN--QVYQ	656	
20		Query: 660	YPMFFDTHLPINKIAVLLKKDSIGNVLMT--SPETYSIKTGHQAQIT 707	
		YP W++D +P I K+KK+ + W S TY++ + I		
	Sbjct: 657	YPTWYDVSVFAGTTTIQKFKIKNG-NTITWEGSNHTTVPSSSTCTVI 705		
An alignment of the GAS and GBS proteins is shown below:				
25	Identities = 112/509 (22%), Positives = 193/509 (37%), Gaps = 103/509 (20%)			
	Query: 18	GNHWKLEESISD--LKKIGISKIWAFFAFKGTSSDQV-----GYGVYDLFLGFEFD	67	
		G W+ + I D L +GIS IW+ + S D GY D F +		
	Sbjct: 79	GGCWQGI IAKDGYLTWGISAIWISSPVENIDSIDFSNGSAAYHGYWAKDFFKTNQH-	137	
30		Query: 68	QNGTIRTKYGRKEEYLKLIKSLKANGIKPFADIVLNHKANGDHKEFKQIKVNPENQGEA	127
		+G + ++ +L+K + IK D NH + + +		
	Sbjct: 138	-----PQTRADPQLVKVARGHHIKVIDFAPFHTSTAEK-----	173	
35		Query: 128	LSEPSIESGWTGDFPGRQGEYNDFFWHYHFTGLDYDAQNNETDIFMIWENKSWADD	187
		G F Y + K G D K+ + +++ W D		
	Sbjct: 174	-----GTTFKEDGALYKNGK-----LVGKFSDDKDK-----IFNHESWTFPS	210	
40		Query: 188	LIDDE--NMFNYLWINDIDFKHPZVLKQWAKWFIETQIGRPLDAVKHDSYFIQ	245
		++ + + N+I+ K + +K D KW G+G R+DAVKH+ + +		
	Sbjct: 211	TVNSIYHSYGLADLANINPKVDQYMKDAID--KWL--DLGVDGIRVDVAGMSQWQK	266	
45		Query: 246	TFINDIRTKIKPLEVFGYWKSDQTSMDYLEATOPOPSLVDVTLRNHFFDASHQ--NRD	304
		+++ I K ++ VFG W S T D + T F + L F A Q A		
	Sbjct: 267	NMLSHIYE--KRNVPVFGG-WFSGHTD--DLYDMTTFANNSGMLLDPRFANAIQVLYG	321	
50		Query: 305	FMRTPFDOSLVIDNDEYA-----VTFVENHDTQSQALESRVEDWFKPLAYGLILR	357
		F T+ D V++N+ VTF++NHID + + + AY L LL		
	Sbjct: 322	FSTFTMRDFYKLENRDQVINEVTDQVTFIDNHMERFATKVMNQTAVNQ--AYAL-LIT	379	
55		Query: 358	QQQTFCLPYGDYIGQGE-----PQQPSFK-----EVIDQWAEIR--QNVVPGQVD	402
		+G P ++YG G+ PSF +VI K+A LR Q +G		
	Sbjct: 360	SRGVNIIYGTBQYATGDKDPNNGGDMPSFNKESQAYKVISKLAPLRKQNALAYGTTBQ	439	
60		Query: 403	YPTHSCNIGWTCLDGEHNSCLAVLITNDQD--GWKHEVGEIYAGCTFPDYLGNC--BQ	458
		+ + + + + + + +A+V N DQ G+ ++D L Q		
	Sbjct: 440	RWISDHVL---VPERKFGNHVALVAINRDQTNQYTTITAKTALPQNSYKDKLGLGQ	495	
	Query: 459	EVVIGDQW-GDFLVSASISAMVKEIE	486	
		E+++G DG F + + ++ W + E+		
	Sbjct: 496	ELIVGADGTISSPFGAGQVAVWVYEGED	524	

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 375

A DNA sequence (GBSx0406) was identified in *S. agalactiae* <SEQ ID 1225> which encodes the amino acid sequence <SEQ ID 1226>. This protein is predicted to be catabolite control protein A. Analysis of this protein sequence reveals the following:

```

5      Possible site: 29
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.2154 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9707> which encodes amino acid sequence <SEQ ID 9708> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database:

```

>GP:BA088121 GB:AB028599 catabolite control protein A [Streptococcus
      bovis] (ver 3)
      Identities = 304/332 (91%), Positives = 320/332 (95%)

20 Query: 1 MNTDDTTITYDVAREAGVSMATVSEVVNKNKNKENTRKVLEVIDRLDLYRPNVAVGLA 60
      MNTDDTTITYDVAREAGVSMATVSEVVNKNKNKENTRKVLEVIDRLDLYRPNVAVGLA
      Sbjct: 1 MNTDDTTITYDVAREAGVSMATVSEVVNKNKNKENTRKVLEVIDRLDLYRPNVAVGLA 60

25 Query: 61 SKKTTTGVVWIPNIANSYFSLARGIDDIAMKYKNIVLASSDEDDKEVNVNTLFAKQ 120
      SKKTTTGVVWIPNIANSYFSLA+GIDDIAMKYKNIVLASSDEDDKEVNVNTLFAKQ
      Sbjct: 61 SKKTTTGVVWIPNIANSYFSLARGIDDIAMKYKNIVLASSDEDDKEVNVNTLFAKQ 120

30 Query: 121 VDGIIIFMHHLTETKIRAEFSRRTFVVLAGTVLSEHQLPSWNIDYKAAVDVIDLAKNN 180
      VDGIIIFMHHLTETKIRAEFSRRTFVVLAGTVLSEHQLPSWNIDYKAAVDVIDLAKNN
      Sbjct: 121 VDGIIIFMHHLTETKIRAEFSRRTFVVLAGTVLSEHQLPSWNIDYKAAVDVIDLAKNN 180

35 Query: 181 KDIAFVSGPLIDDINGKVRLAGYKSELKNGNLNFKESGLVFEANYRYAGFALAQRVINAG 240
      KDIAFVSGPLIDDINGKVRLAGYKSEL+KN L+FKESGLVFEANY Y +G+ LAQRV+N+G
      Sbjct: 181 KDIAFVSGPLIDDINGKVRLAGYKSELKNGNLNFKESGLVFEANYRYAGFALAQRVINAG 240

40 Query: 241 ATAAYVAEDELAAGLLNLGLFAGKRVPEDEFTITENDSPIAQYTRPNLTSSQPVYDLGA 300
      ATAAYVAEDELAAGLLNLGLF AGK+VPEDFRI+TENDSPI YTRPNL+SSSQPVYDLGA
      Sbjct: 241 ATAAYVAEDELAAGLLNLGLFAGKRVPEDEFTITENDSPI+TTRPNLSSSQPVYDLGA 300

40 Query: 301 VSMRMLTKIMKEELEKEKEIILNHGILKAGTT 332
      VSMRMLTKIM+KEELEKEEI+LNHG+ RGTT
      Sbjct: 301 VSMRMLTKIMKEELEKEEIILNHGILKAGTT 332

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1227> which encodes the amino acid sequence <SEQ ID 1228>. Analysis of this protein sequence reveals the following:

```

45      Possible site: 29
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
50      bacterial cytoplasm --- Certainty=0.2154 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

55      Identities = 307/332 (92%), Positives = 320/332 (95%)

      Query: 1 MNTDDTTITYDVAREAGVSMATVSEVVNKNKNKENTRKVLEVIDRLDLYRPNVAVGLA 60
      MNTDD +TITYDVAREAGVSMATVSEVVNKNKNKENTRKVLEVIDRLDLYRPNVAVGLA
      Sbjct: 1 MNTDDPTITYDVAREAGVSMATVSEVVNKNKNKENTRKVLEVIDRLDLYRPNVAVGLA 60

```


-479-

Query: 61 SKKTTTVGVVFNANSYFSLAIGLIDDIAAMYKYNIVLASSDEDDDEKVNVTFLPAKQ 120
 SKKTTTVGVVFNANSYFSLA+GIDDAIAMYKYNIVLASSDEDDDEKVNVTFLPAKQ
 Sbjct: 61 SKKTTTVGVVFNANSYFSLAIGLIDDIAAMYKYNIVLASSDEDDDEKVNVTFLPAKQ 120

5 Query: 121 VDGIIFMGHHLLTKIRASFSSRRTPIVLAGTVDLHQPLFSVMIDYXAAVVDVILAGNH 180
 VDGIIFMGHHLLTKIRASFSSRRTPI+VLAGTVDL+HQLPSVMIDY+AA +V+DLIA NH
 Sbjct: 121 VDGIIFMGHHLLTKIRASFSSRRTPIVLAGTVDLHQPLFSVMIDYXAAVVDVILAGNH 180

10 Query: 181 KDIASFVSGPLIDDINGKVRLAGYKGLKKNLNFKEGLVFEANYRVAAGPALAQRVING 240
 K IASFVSGPLIDDINGKVRLAGYKGLK N L+FEGLVFEANY Y BGF LAQRVIN+G
 Sbjct: 181 KDIASFVSGPLIDDINGKVRLAGYKGLKKNLNFKEGLVFEANYRVAAGPALAQRVING 240

15 Query: 241 ATAAVVAEDELAAAGLLNGLFEAGKRVPEDFRIITSNDSPIAQYTRNLTSISQPVYDLGA 300
 ATAAVVAEDELAAAGLLNGLFEAGKRVPEDFRIITSNDSGP+ QYTRNL+LSISQPVYDLGA
 Sbjct: 241 ATAAVVAEDELAAAGLLNGLFEAGKRVPEDFRIITSNDSPPVQYTRNLTSISQPVYDLGA 300

20 Query: 301 VSMFMTKIMHKEBLEKEKIVLNHGIKRGTT 332
 VSMFMTKIM+KEBLEKEKI+LNHGI KRGTT
 Sbjct: 301 VSMFMTKIMHKEBLEKEKIVLNHGIKRGTT 332

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 376

25 A DNA sequence (GBSx0407) was identified in *S. agalactiae* <SEQ ID 1229> which encodes the amino acid sequence <SEQ ID 1230>. This protein is predicted to be PepQ (pepQ-2). Analysis of this protein sequence reveals the following:

Possible site: 22
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1118 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

35

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAc46293 GB:AF014460 PepQ [Streptococcus mutans]
 Identities = 257/359 (71%), Positives = 304/359 (84%)

40 Query: 1 MSKLNRRHLLHSQARLAVFSDPTVNYLTGFCDFPHERQMFVYEDRDPILEVPAL 60
 MSKL +I L E AV SDPV++NYLTGF+ DPHSR MFLF++ D++ +LF+P L+
 Sbjct: 1 MSKLAQIVQKLKKQCIERAVLSDPVSINYLTFPSDPHERLNFLLFPADQETLLFLPELD 60

45 Query: 61 VSRKQSGVEFFVPGYIDSENFQWQIAENLFGPSVSKVLARFDNLNVTKPQGLQTVFDFGH 120
 RAK + V GY+D ENP +KI +LP + SK+ EFDNLNVTKF+GL+T+F G F
 Sbjct: 61 ALRKAQSLDSVYGYLDFENPLEKIKTLPEPTNYSKIALEFDNLNVTKFKGLRTIPSGQF 120

50 Query: 121 ENLTPYIKMRLIKSRDEIEKMLVAGSFADFAVQVGFNINSLNNTETDIIAQIEFPMKKQ 180
 NLT P I MRLIKE DEI+K+L+AGR ADKAVQ+GFD+ISLN TS/DIIAQIEFPMKK
 Sbjct: 121 TNLTPLINRMRLIKSADIEIKLLTAGRLADFAVQIGPDSISLNATETDIIAQIEFPMKKL 180

55 Query: 181 GINKMSFTMVLTVGNNAANPHGIPGINKIENNALLLFDLGVTLSYTSMTNRTVAVGKPD 240
 G++KMSF+TMVLTV+NAANPHG+P ++KIENN LLLFDLGVE+ GY SDMTNRTVAVG+PD
 Sbjct: 181 GVDKMSFTMVLTVGNNAANPHGLPASHKIKENNLALLFDLGVESTGYSDMTNRTVAVGKPD 240

60 Query: 241 QFKKDIYHLCEAHQAIDFIPKGVLADEVDAARNVIERKAGYGYFPHRLGRLGMDVH 300
 QFKKDIY++CLSEA A+DFIKPGV A++VDAAR+VIERKAGY YPHRLGRLG+GM +H
 Sbjct: 241 QFKKDIYNICLAQLTALDFIPKGVSAQVDANARSVIERKAGYGYFPHRLGRLGIGLH 300

65 Query: 301 EFPISAGNDMEIQBGMCFSEVPGIYIPDKVGVRIEDGYSYVTKRGFEVFTKTPKELLYF 359
 EFPISAGNDM ++BGMCFSEVPGIYIP+KVGVRIEDG+VTK GREVFT+TPKELLYF

Subjct: 301 EFPSIMAGNDMLLEBGMCFSPVEPGIYIPEKGVRIEDCGHVTONGFEVFTQTPKELLYF 359

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1231> which encodes the amino acid sequence <SEQ ID 1232>. Analysis of this protein sequence reveals the following:

5 Possible site: 58
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.90 Transmembrane 42 - 58 (42 - 59)

 10 ----- Final Results -----
 bacterial membrane --- Certainty=0.1362 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

15 >GP.AAC46293 GB:AF014460 PepQ [Streptococcus mutans]
 Identities = 264/359 (73%), Positives = 304/359 (84%)

 20 Query: 1 MTKLDQIRLYLDQGRALAI FSDPVTINYLTFGFCDPHERQQLFLVYHDLAPVLVFPVPALE 60
 N+KL QI L ++G E A+ SDPV+INYLTFG+ DPER +FLF++ D +LF+P L+
 Sbjct: 1 MSKLAQIVQKRLKQIGBAVLSDPVSINYLTFGYSDPHERINLFLFLADQETLLFPLG 60

 25 Query: 61 VARASQAISFPFPGYVDSENFWEIKAVLNTAAKTIYAEFDHLVNFKPHGLQITFSGQF 120
 RA + V GY+D ENP ERIK +L+ T I EPD+INV KF GL+TIFSGQF
 Sbjct: 61 ALRAKSILOISVTGVLDFENPLEKIKITLLPKTINYKISIALEFIMLVNTRKFGLETIFSGQF 120

 30 Query: 121 NNLTPYVQGMRLVKSADENQMIAGQFADKAVQVGFNDISLQATETDIAQIEFEMKQ 180
 NLTP + NRL+KSADEI K++LAG+ ADKAVQ+QFD+ ISL+RTETD+IAQIEFEMK
 Sbjct: 121 TNLTPLINRHLIKSADIEKQIKLLIAGELADKAVQIGFDSISLQATETDIAQIEFEMK 180

 35 Query: 181 GIHKMSFDTWLVTGNNANPHGIRGTINIKENALLFDLGVETLGYTSDMTRTVAVGQPD 240
 G+ KMSF+IMVLTS+NAANPHG+P ++ IENN LLLFDLGV+ GY SDMTRTVAVGQPD
 Sbjct: 181 GVDKMSFETWLTGNNANPHGLPASHKIENNHLILLFDLGVESTGYSDMTRTVAVGQPD 240

 40 Query: 241 QPKIDINYLCLEAQLAIDFIKPGVTAQVDAARVIEKAGYGVEYFNHRLGHGIMDVH 300
 QPK DIYN+CLEAQL A+DPFKGV+AAQVDAAR VIEKAGY+YFNHRLGHGIM +E
 Sbjct: 241 QPKKIDINYLCLEAQLTALDFIKPGVSAQVDAARVIEKAGYGDYFNHRLGHGIMGLH 300

 Query: 301 EFPSIMAGNDMLLEBGMCFSPVEPGIYIPKGVRIEDCGHVTONGFEVFTQTPKELLYF 359
 EFPSIMAGND++LEBGMCFSPVEPGIYIP KGVRIEDCGHVTONGFEVFT TPKELLYF
 45 Sbjct: 301 EFPSIMAGNDMLLEBGMCFSPVEPGIYIPEKGVRIEDCGHVTONGFEVFTQTPKELLYF 359

An alignment of the GAS and GBS proteins is shown below:

Identities = 288/361 (79%), Positives = 325/361 (89%)

45 Query: 1 MSKLNRIKHLHSVQRLAVPSDPTVTINYLTFGFCDPHERQQLFLVYHDLAPVLVFPVPALE 60
 N+KL++IR +L ASLA+FSDPVT+NYLTFGFCDPHERQ+FLVY D P+L+VPALE
 Sbjct: 1 MTKLDQIRLYLDQGRALAI FSDPVTINYLTFGFCDPHERQQLFLVYHDLAPVLVFPVPALE 60

 50 Query: 61 VSRKQSVFPFPGYVDSENFQKIASNMLPSVSKVLAEFTNLANVTRKQGLQCVVDGHP 120
 V+RA Q++ FVPGY+DSENF+KI +L+ + + AEFD+LAV KF GLQ+P G F
 Sbjct: 61 VARASQAISFPFPGYVDSENFWEIKAVLNTAAKTIYAEFDHLVNFKPHGLQITFSGQF 120

 55 Query: 121 ENLTPYVQGMRLVKSADENQMIAGQFADKAVQVGFNDISLQATETDIAQIEFEMKQ 180
 NLTPY+Q MRL+KS DEI KM++AG+ADKAVQVGFNDISL+ TETD+IAQIEFEMKQ
 Sbjct: 121 NNLTPYVQGMRLVKSADENQMIAGQFADKAVQVGFNDISLQATETDIAQIEFEMKQ 180

 60 Query: 181 GIHKMSFDTWLVTGNNANPHGIRGTINIKENALLFDLGVETLGYTSDMTRTVAVGQPD 240
 GI+KMSFDTWLTGNNANPHGIRGTINIKENALLFDLGVETLGYTSDMTRTVAVG+P
 Sbjct: 181 GIHKMSFDTWLVTGNNANPHGIRGTINIKENALLFDLGVETLGYTSDMTRTVAVGQPD 240

 Query: 241 QPKIDINYLCLEAQAIDFIKPGVTAQVDAARVIEKAGYGVEYFNHRLGHGIMDVH 300
 QPK DIY+LCLEA AAIDPIKGV A++VDAAR VIEKAGY+YFNHRLGHGIMDVH
 Sbjct: 241 QPKIDINYLCLEAQAIDFIKPGVTAQVDAARVIEKAGYGVEYFNHRLGHGIMDVH 300

-481-

Query: 301 EFPPSIMGNDMEIQGMCFSVEPGIYIPDKGVRIEDCGYVTKGPKVFTKTPKELLYFEG 361
 EFPSIMGND+ ++EGMCFSVEPGIYIP DKGVRIEDCG+VTK GPKVFT TPKELLYFEG
 Sbjct: 301 EFPSIMGNDLVLEEGMCFSVEPGIYIPDKGVRIEDCGHVTKNGPSVPTHPTKELLYFEG 361

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 377

A DNA sequence (GBSx0408) was identified in *S.galactiae* <SEQ ID 1233> which encodes the amino acid sequence <SEQ ID 1234>. Analysis of this protein sequence reveals the following:

10 Possible site: 14
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3629(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 378

A DNA sequence (GBSx0409) was identified in *S.galactiae* <SEQ ID 1235> which encodes the amino acid sequence <SEQ ID 1236>. This protein is predicted to be beta-hexosamidase A precursor. Analysis of this protein sequence reveals the following:

25 Possible site: 47
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3279(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

35 >GP:CAB11942 GB:Z99104 alternate gene name: yzba-similar to
 beta-hexosaminidase [Bacillus subtilis]
 Identities = 151/602 (25%), Positives = 268/602 (44%), Gaps = 69/602 (11%)

Query: 26 INEMTLDKIKQLF-----FNMGASRSEKYLTDVLDRIYHIAVRYNRGSSELYDQNL- 78
 +N M+LDEK+GQ+ + S + LT + D +Y G ++ +N+
 Sbjct: 39 VNRMSLDEKLGMLMPDFGNWQKGESEPPALITMNDDEVASLVKKIQFGGII-LFAEHW 97

Query: 79 -----ILQTKSKLMLIAANTEAGGDAVTLGTGKVEDIKVAATNPDKYAYEMG 127
 + K+P++++ + E G + +GT + + A AY+ G
 Sbjct: 98 TTKQTIVQLTIDYQKASPKI PLMSLDGEGSIVYRLGEGTNFPGNMGLGAARSRINAYQTG 157

Query: 128 RIAGGEASAVGCNASFSPVVDLTNNRNPITASRNWGANVDQILSLSEKYMKGIMQYINIV 187
 I G E SA+G N FSP+VD+ N NP+I R++ +N + L MFG+ + +I
 Sbjct: 158 STIGKEISALGINTDFSPVVDINNNPDPVIGVRSPSSNRELTSRLGLYTMGLGQDIA 217

50 Query: 188 PPAKIFPGDGIDRDEHLFSASNPMSEKEMSTPGRIYGLADAGLPGWAGHITHLHVE 247
 KHFFG G + D H +E + + DAG VM H+ P +
 Sbjct: 218 SALKHIFPGHGDITVDSHYGLPLVSHOQERLREVELYPPQKALDAGADMVMTARVQFPAPF 277

-482-

Query: 248 KEMHPR--DLDMLPASLNKTLLELLAGSLGYNGAIVTASHVGVMTASMRRLDLP 305
 + + D ++PA+L+K ++ LLR E+G+NG IVTDA +M + + + +
 Sbjct: 278 DTTYSKLDGSDILVPATLSKKVMTGLLRQEMGFNGVIVTDALNMKAADHPQGEAVVM 337

5 Query: 306 ATRAGCDLWLF--FNDDPD-----IQWKKRGYKGLITRERLHDALEPTTLGLKRLG 356
 A++AG D+ L E+ IQ +KE + G + E+++++ R + LK K G
 Sbjct: 338 AVKGVGDALMPASVTSIKSQKQFARVIGALAEAVKNGDIPEQQINNSVRIISLTKRG 397

10 Query: 357 LHNYSQRQSLFMPK--DKAMALINTLSQKLIADRVADKAVTLVKDKQKDIFFVNPERYH 415
 + Y R + K KA ++ + K ++AKVTH+K+Q + D P+
 Sbjct: 398 M--YPARNSDSTKEKIKAKKCTVGSQHLKAEKKLAKKAVTVLKNQHTL--PPKPKKGR 454

Query: 416 ILLNVREGYKGGFGAMIAQHKQASDYMKR-----LLEARGHEVTVWRSTERTIKLPQ 469
 IL+V + A +Q D+K L V+++ E+ +K
 15 Sbjct: 455 ILLV-----APFEQTASIEQTTHDLTKKKKIKPVSLKMNFAQVQKTEHEKQVK-- 505

Query: 470 EERRAALANVYFAK-QFIANLTHEYDLINLVVDVNRGGTQRLIWPAAKGTDPQPFVYH 528
 E I Y K P+ N D+I+ D + + ++P A + H
 20 Sbjct: 506 -EADYIITGSYVVKNDPVN-----DGVID--DTISDSSKMATVFPRA---VMAALQHN 554

Query: 529 IPISTIVSQHAPALADMPQVQYIINAYD-----GLPSTISAVVAKLAGESLEPTVSP 580
 P++S+++ A+ + I Y L I A V +G++ G P
 Sbjct: 555 KPVVLSLNFYDAANFEENALIAVYGFQYANGRYLQPNIDAGVMAIPQAPKGTLP 614

25 Query: 581 VD 582
 VD
 Sbjct: 615 VD 616

No corresponding DNA sequence was identified in *S.pyogenes*.

- 30 A related GBS gene <SEQ ID 8565> and protein <SEQ ID 8566> were also identified. Analysis of this protein sequence reveals the following homology to a lipoprotein, with homology with the following sequences in the databases:

29.5/52.3% over 422aa

- Bacillus subtilis
- 35 EGAD|20114| hypothetical 70.6 kd protein in feua 5'region precursor Insert characterized
 SP|P40406|YBBD_BACSU HYPOTHETICAL 70.6 KDA LIPOPROTEIN IN FEUA-SIGW INTERGENIC REGION
 PRECURSOR (ORF1). Insert
 characterized
 GP|1944006|dbj|BAAI9499.1||AB002150 YbbD Insert characterized
- 40 GP|438455|gb|AAA64351.1||L19954 possible N-terminal signal sequence; mature protein may
 be membrane-anchored and start at Cys-17. 17.5% identity
 over 354-aa overlap with Candida pelliculosa beta-glucosidase.; putative Insert
 characterized
 GP|2632433|emb Insert characterized
- 45 ORF00431(367 - 1557 of 2388)
 EGAD|20114|RS0166(36 - 458 of 642) hypothetical 70.6 kd protein in feua 5'region precursor
 {Bacillus subtilis} SP|P40406|YBBD_BACSU HYPOTHETICAL 70.6 KDA LIPOPROTEIN IN FEUA-SIGW
 INTERGENIC REGION PRECURSOR (ORF1). GP|1944006|dbj|BAAI9499.1||AB002150 YbbD {Bacillus
 50 subtilis} GP|438455|gb|AAA64351.1||L19954 possible N-terminal signal sequence; mature
 protein may be membrane-anchored and start at Cys-17. 17.5% identity over 354-aa overlap
 with Candida pelliculosa beta-glucosidase.; putative {Bacillus subtilis} GP|2632433|emb
 %Match = 9.6
 %Identity = 29.5 %Similarity = 52.2
 55 Matchee = 119 Mismatches = 183 Conservative Sub.s = 92
- | | | | | | | | |
|---|-----|-----|-----|-----|-----|-----|-----|
| 114 | 144 | 174 | 204 | 234 | 264 | 294 | 324 |
| LMVGSLSGLAAAEQNGIAFYFVIVGKIVKSWELLREDIGRAFAGQEQEQRKBSINTFWANLDN**KG*AMTHINDLT | | | | | | | |
- 60 MRPFVPLILSAVLELSCFFGA
 10 20
- | | | | | | | |
|---|-----|-----|-----|-----|-----|-----|
| 354 | 384 | 414 | 426 | 456 | 486 | 528 |
| KKPFNLNQALAEIWEKTIENMTLDEKIGQLFF-----NMGARSSEELYTLVLRKYHIAAVNRNGS-----SSRTYDQ | | | | | | |

-483-

```

      :      :      :      :      :      :      :      :      :      :      :      :      :      :      :
      RQTEASASKRAIDANQIVNRMSLDEKLGSLMFDPRNQKEGSSPQALTRNDEVASLVKKYQPGGIILFAYENVKTTTQ
      40      50      60      70      80      90      100

5      543      573      603      633      663      693      723      753
      NLIL-----QTSKSLPMLIAANTEAGGCGAVITDGTKVGEIKVAATNDPKYAYEMGRLAGMEASAVGCGASPSFIVDLTR
      :      :      :      :      :      :      :      :      :      :      :      :      :      :      :
      TVQLTDDTYQKASPKIPMLSLIDQBGGLIVRLGSLTSPFGMMALGAARSINATYQTGSIIGKLSALGINTDFSPFVVDINN
      120      130      140      150      160      170      180

10      783      813      843      873      903      933      963      993
      NWRNFIIASRNWGNVDQIIISLSKEYMKGIMQYINIVFAKHFFPGDGIDERDHHLSFASNPMSEKEMASTFGRIYIGELADA
      :      :      :      :      :      :      :      :      :      :      :      :      :      :      :
      NEDNPFVIGVSPSSNRELTSRLGLYTMKGLQRQDIASALKHFFPGHSDTVDSHYGLPLVSHQERLREVLFFPQKRAIDA
      200      210      220      230      240      250      260

15      1023      1053      1080      1107      1137      1167      1197      1227
      GLFGVMAGHIEHFNVEKEMHFER-DLDDML-PASLNKTLIDELLARSELGYNGAIVTDASHVVGVTASGNARRDLLFTALEA
      :      :      :      :      :      :      :      :      :      :      :      :      :      :      :
      GADNVMVTAHVQFPAFDDTYKSLDSSDILVPATLSKUKVMGLLRQEMFNGVITDALNNKAIADHSQEAIVMRAVKA
      280      290      300      310      320      330      340

20      1290      1320      1350      1380      1410      1437
      GCDLFLP-----PNDPDE---DIQNMKSGYERKILTEERLHDLRRTLGLKAKGLHNYBGRQELTFPK-DKAWALIN
      :      :      :      :      :      :      :      :      :      :      :      :      :      :      :
      GVDIALMPSVTSLKBEQFARVITQALKZAVKNGDIPQQTINNSVERIISLAKIKRGM--YEARNSDSTKEIAIAKXKIVG
      360      370      380      390      400      410

25      1467      1497      1527      1557      1587      1617      1647      1677
      TLESQKIADEVADKAVTIVKDKQKDIFFPNFERYRHILVNVEGYKGGPGMIAGNKQASDYMKELLEARGHEVTVNES
      :      :      :      :      :      :      :      :      :      :      :      :      :      :      :
      SKQLKAKKELAKAVTVLAKNEQ-HTLPFPKPKGSRILIVAPYERQTASIRQTIHDLIRKKIKPKVSLSKOMIFASGVQFKT
      430      440      450      460      470      480      490

```

SEQ ID 1236 (GBS50) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 11 (lane 8; MW 69.2kDa).

GBS50-His was purified as shown in Figure 192, lane 5.

The GBS50-His fusion product was purified (Figure 192, lane 5) and used to immunise mice. The resulting antiserum was used for FACS (Figure 264), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 379

A DNA sequence (GBSx0410) was identified in *S. agalactiae* <SEQ ID 1237> which encodes the amino acid sequence <SEQ ID 1238>. Analysis of this protein sequence reveals the following:

```

Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2266 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 380

A DNA sequence (GBSx0411) was identified in *S.galactiae* <SEQ ID 1239> which encodes the amino acid sequence <SEQ ID 1240>. Analysis of this protein sequence reveals the following:

Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2279 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9705> which encodes amino acid sequence <SEQ ID 9706> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>P:AAC21726 GB:U32690 oxidoreductase [Haemophilus influenzae Rd]
Identities = 197/271 (72%), Positives = 229/271 (83%)

Query: 26 NKVVVITAGGVLGYMAKEFAKAGAKVALLDLNQEAAQTFADEIVBEGGIKAYKANVL 85
NK+++ITGAGGVLC ++AK+ A A +ALLDLN ERA A EI + GG AKAYK NVL
Sbjct: 15 NKLIIITAGGVLCSFLAKQLAYTKANIALLDLNFPAADKVAKEINQSGGKAKAYKTNVL 74

Query: 86 SKENLEEVHQAVLEDLGPDIILVNGAGGNNPKATTIDNEPHELDLPSTTKTFPELDEAGIS 145
EN++EV + D G DIL+NGAGGNNPKATTIDNEFH+ DL T+TF+LD++GI
Sbjct: 75 ELENKEVFNQIBTDPTGCDILVNGAGGNNPKATTIDNEPHELDLPSTTKTFPELDEAGIS 134

Query: 146 FVFNLNVLGTLTLPQTQVFAQDMVREGGANIINISSMNAFTLTLPKIPAYSGAKAAISNFTQW 205
FVFNLNVLG+LLPQTQVFA+DM+G+GANIINISSMNAFTLTLPKIPAYSGAKAAISNFTQW
Sbjct: 135 FVFNLNVLGSLTLPQTQVFAKDMLGKQGANIINISSMNAFTLTLPKIPAYSGAKAAISNFTQW 194

Query: 206 LAVHPSKVQIRCNALAPGFLVTNQNRSLLPFTDGGQPTARAEEKILNNTPMGRFGEASELIG 265
LAV+PSKVQIRCNALAPGFLV+NCN +LLF +G+PT RA KIL NTPMGRFGE+ EI+G
Sbjct: 195 LAVYPSKVQIRCNALAPGFLVSNQNLALLFDTSKPTDRANKILNTPMGRFGESEELG 254

Query: 266 GLFFLADEKSSSPFNGVVLPIDGQFAAYSQV 296
L FL DE S+PVGVLPL+DGGF+AYSQV
Sbjct: 255 ALLFLIDENYSAPFNGVVLPIVDGQFSAYSQV 285

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1241> which encodes the amino acid sequence <SEQ ID 1242>. Analysis of this protein sequence reveals the following:

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0358 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 77/279 (27%), Positives = 125/279 (44%), Gaps = 19/279 (6%)

Query: 18 MSKTTITPINKVVVITAGGVLGYMAKEFAKAGAKVALLDLNQEAAQTFADEIVBEGGIA 77
M + K+ +ITGA + +AK +A+AGA + D+ QE E G A
Sbjct: 1 MSNPSLQCKIALITGASYGIGPHLAKAYAGATTVPNDIKQELVDKGLAAYRELIGEA 60

Query: 78 KAYKANVLSKENLEEVHQAVLEDLGPDIILVNGAGGNNPKATTIDNEPHELDLPSTTKTF 137

-485-

Y +V + +++++ +++++ DILVN AG
 Sbjct: 61 HGFVCDVTEAGIQQVSGQTEDEVGALDILVNNAG-----IIRTPML 103
 5 Query: 138 ELDEAGISVFNLYNLGTLTLPQVFAQDMVGRGANLINISSMNAFTPLTKIPAYSQAKA 197
 E+ V +++ + ++ + + LIMI SM + + AY+ AK
 Sbjct: 104 EMAAEDFRQVIDDLNAFTVSKAVLPSMAKGHGKTIINICSMSELGRSTVENYAAAG 163
 Query: 198 AINSFTQWLAVHPSKVGIRCMALAPGLVITVQNRSLRTE-DQQTARAKNIANTPMGR 256
 + T+ +A F + I+CN I RG++ T Q L + DG + I+ TV R
 10 Sbjct: 164 GLKMLTKNIASEFGEANTCQNGIQPGYIATPCTAFALRGRQAGSRHPFDQFIATKTPAAR 223
 Query: 257 PQSASELIGSLFFLADEKSSFPNGVILPIDGGFAAYSG 295
 +G +L G FLA + +S+PVNG +L +DGG AY G
 15 Sbjct: 224 MGTTEDLACPAVFLASD-ASNFPVNGHILYVDGGLIATIG 261

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 381

A DNA sequence (GBS0412) was identified in *S. galactiae* <SEQ ID 1243> which encodes the amino acid sequence <SEQ ID 1244>. This protein is predicted to be D-mannonate dehydrogenase (uxuA). Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3188 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04425 (GB:AP001509 D-mannonate dehydrogenase [Bacillus halodurans])
 Identities = 202/343 (58%), Positives = 261/343 (75%)

Query: 1 MEMSFRMYGDEDPVTLENIGQIPDMKGIIVTAIYDVFGVWMSRRIQQUKEKVEAAGLKI 60
 M ++ RM+G D V LE I QIP MGGIV+AIYDV VG VW +E+I LK +E GL +
 Sbjct: 1 MRLTMRWFGSGEDKVKLEYIKQIPMGKGIIVTAIYDVAVGGWVPEKILALNNIERHGLTL 60

Query: 61 SVIESVVPVHEDIKLRGPTDLIDNYIQTVNLAEGIDITCYNMFVFWQTRIDLAYQY 120
 VIESVVPVHEDIKLG+PTRD I+NY QT+++LA GIDT+CYNMFVFWQTR+ L ++
 40 Sbjct: 61 DVIESVVPVHEDIKLGKPTRDRYIENYKQTLRLHAECDITVTCYNMFVFWQTRSLQDFKL 120

Query: 121 PDGSTALIFDVTYKMDPFGVNGELSLPGWDASYSKREKAIMDQYAEIDEEKLWNLTYF 180
 DGS ALI++E V + +P+GEL LPQWD SY E +K ++ AY +I EE LW++ILTYF
 45 Sbjct: 121 EDGSEALIEEDVTSRTNPLSGLELPGWDTSYENSLGVLQAYKKISSEDLWDHLTYF 180

Query: 181 IKRIIFAEAAVGVMAIHFDPPPSYIFGLPRIITGLEATYERFVKLYDSKNGITLVGVS 240
 ++ I+P A+ VG+KMAIHFDPP+SIIFGLPRI+T +ER + LYDS ++GIT+C GS
 Sbjct: 181 VQAIMFVADEVGKMAIHFDPPPSYIFGLPRIIVINKANLERLLSYDSPNHGIMTCSQSL 240

Query: 241 ASDPQNDVLEISRRAFHLDRVFNVHARNIKLGDGSKFKBSAHPSEYGSIDMYEVIKLCHS 300
 ++ ND+ E+ R R++F HARNIK +SF+RSAR SR GS++M ++K H+
 Sbjct: 241 GANNEANDLPMIRHFGGGRIFHFAHARNIKTGPRSPQESAHLSAGSVNVM/LMKAYHD 300

Query: 301 PGFEGAIRFDHGRMIDGTRGPGYGLYRGLGATYVGLYEAV 343
 GP G +RFDHGRMIDG GRPGYGLYRGLGATY+G++RAV
 55 Sbjct: 301 IGPFGPLRFDHGRMIDGGRPGYGLYRGLGATYVGLYEAV 343

No corresponding DNA sequence was identified in *S. pyogenes*.

-486-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 382

A DNA sequence (GBSx0413) was identified in *S.galactiae* <SEQ ID 1245> which encodes the amino acid sequence <SEQ ID 1246>. Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2447(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 383

A DNA sequence (GBSx0414) was identified in *S.galactiae* <SEQ ID 1247> which encodes the amino acid sequence <SEQ ID 1248>. This protein is predicted to be uronate isomerase. Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3066(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04424 GB:AP001509 uronate isomerase [Bacillus halodurans]
Identities = 215/465 (46%), Positives = 294/465 (62%), Gaps = 7/465 (1%)

Query: 3 FNTETFMKLNQAAIQLYEE-VKROPIDYVCHLPKDI FDIHIFDNI VDLWGGDHYGNR 61
F +E F+L N+ +LY K PI DYHCHL P++I+Z+ F+N+ WLGGDHYGNR
Sbjct: 4 FLSEDFLLNNEYDRELYYTFAKNMPICDYHCHLSQRIWENKPFENMTKANLGGDHYGNR 63

Query: 62 LMRANGISEAEITGASNLSEKFKAFARTLERAYGNPVYHNSAMELNKVFVGNELTSENA 121
MR NG+ E ITG A + EK F A+A+T+ + GNP+YHN+ MELK F ++ L E+N
Sbjct: 64 AMRLNGVREEFTTGAPDKKFLAWAKTVPKTIQNPVYHNSAMELNKVFVGNELTSENA 123

Query: 122 KEIYHRLNHLKEHKI SPRLIADSKVMFICTTIDHPLDTLEWHKGLAADSFKTVVAPTF 161
E ++ N L++ +PR LJ S V IGITD P D+L +H+KL AD++F V PTF
Sbjct: 124 ENVDACNRLIQEAPTFPRALIERSNVRAIGTDDPTDSELLYHQIQADDTFHVKVITP 183

Query: 182 RFDEAF-IEHRHFVDPIITKLGSDITQKEITDFSTFIAMBERIAYFQNGCRASDISFTPI 240
RPD A IR F D++ KL D+T + + F+ A++ER+ +F +GCR+SD TS+
Sbjct: 184 RPDGALKISQDSFDWVAKLSDVTGSELDLDAFLHAKERLTFPDEHGCRSSDHDMTEV 243

Query: 241 VFEQTDSEGLNDLPNKVCSGYIPNQSEISKQATVFMELCRLYKKYGFVTVQVHGLRNN 300
F + +E E +F +K +E K++T + L +Y G+V Q H G +RNN
Sbjct: 244 PFVEVNEQAOHIFPEKRLANGLTKVENEKYTFMTWLKGEYAARGWVQWHLIGVNRN 303

Query: 301 HSTIFEKIGADVGVDLSGD-QVALTVNWRNLLDSLVKKDSLPMTIYNINPAYNIAYANT 359

-487-

+S + KLG D G DS+GD Q+A +LLD L K+ +LPK I Y +NP N +A+
 Sbjct: 304 NSRMLHLKGFDTGFDSTIGDQIAHAT--AKLLLDLRQKALPATILYCVNPNTILASN 361
 Query: 360 LANPANELGVRSYLQFGAGWVFAATKLGWISQNALABQCLANFICMLTDSRSFLSYQ 419
 + NF E GVR +QEG+ WKF D GM Q+ LA G+L+NFICMLTDSRSFLSY
 Sbjct: 362 IGNF--TBGGVRGKVGQFGSANWFNDHIDGMRQUTDLASVGLLENFICMLTDSRSFLSY 419
 Query: 420 RHDYFRRLLCTYLGEWIEBGEVPEYQALGSMNAKDIAVQNAVWF 464
 RHDYFRRLLC +G WI+EG++P D + G + +DI Y N V+YF
 Sbjct: 420 RHDYFRRLLCQLLGSWIKGQLPPMERWQVQIDICTNVDYF 464

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

15 Example 384

A DNA sequence (GBSx0415) was identified in *S.agalactiae* <SEQ ID 1249> which encodes the amino acid sequence <SEQ ID 1250>. This protein is predicted to be 2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate al. Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3883(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9703> which encodes amino acid sequence <SEQ ID 9704> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

30 >GP:AA035160 GB:AE001693 2-dehydro-3-deoxyphosphogluconate
 aldolase/4-hydroxy-2-oxoglutarate aldolase [Thermotoga maritima]
 Identities = 93/199 (46%), Positives = 125/199 (62%), Gaps = 6/199 (3%)
 Query: 37 KNNYFPAVIRGKSSDALSLAKHAILGGIRNIEVTFSTPEASKVTKQLSDDFKNHKEIIV 96
 K + AV+R S E+ A E A GG+ IE+TF+ P+A VIK+LS F K I +
 35 Sbjct: 8 KKHKIVAVLRANSVESAKKALAVFEGGVHLIETFTVPDADTVIKELS--FLKEKGAI 65
 Query: 97 GAGTVMITELAKEAIDAGAKFLVSPHPLSDIANLANENKVYFPGCATATEIIVVARKYK 156
 GAGTV + E ++A+++GA+F+VSPH D +I+ E V+Y FG T TE+V A K
 40 Sbjct: 66 GAGTVTSVEQCKAKVSSGARFIVSPHLEDEISQCKEKGVPYMGVMTPLVKAKMLGH 125
 Query: 157 QIILKFPQGVGPGFIKIDHGPIDVDLMPGGVSVSNVEMRKAGAVAVGVGSALSSKV 216
 I+ILKFPQGVGPGF+K + GP P+V +P+GGV+ NV EN KAG +AVGVGSAL
 Sbjct: 126 TLILKFPGEVVGQFPVKAMKGFPFNKVFPTGVNLDNENCFKAGVLA VGVGSALVKG 185
 45 Query: 217 MTEGYDSVTKIAKQFVSAL 235
 D V + AK FV +
 Sbjct: 186 P----DEVREKAKAFVKRI 200

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1251> which encodes the amino acid sequence <SEQ ID 1252>. Analysis of this protein sequence reveals the following:

Possible site: 29
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1039(Affirmative) < succ>

-488-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 82/204 (40%), Positives = 132/204 (64%)

Query: 32 MLNQLKNNYFFAVIRGKSSSEDALEIAHAILAQIRNIEVTFSTPENSKVIRQLSDDFKNN 91
+L +LK N V+RG+SSB+AL + +I GGI+ IEVT+ P AS+VI QL+ + FK +
Sbjct: 6 ILTKLKANRLVLVVRGSSSEALACSIAIGGKINTIEVTTNPFASEVIGLAERFKD 65

Query: 92 KRILVGAQTVMTLEAKRAIDAGAKFLVSPHFDSDIANLANENKVYYPGCATATEIVVA 151
E+++GAGTV+ A++AI AGA+F+V P+F+ +A + + + Y POC T E+V A
Sbjct: 66 PEVLIGAGTVLDOVTARQAILAGAQFVGPENFRAVALICHRSIPYLPQMTVEVITA 125

Query: 152 RYKQKQIILKLPFGGVVGPFIKIDHGFIPDVLMPGGVSVSNVWNRKAGAVAVGVGA 211
+ ++K+PFG VG FI+ I P+P V++M +GGVS N+ +W AG +C+G
Sbjct: 126 LBSGVDMVKIFPGSTVGIISFIRAIKSLPQVEVMVTGGVSDNLKDLAAGVTVLIGIG 185

Query: 212 LBSKVATEGYDSVTIKARQFVSAL 235
+ + + Y+ +TK A ++ +L
Sbjct: 186 FNQLASQKQYNLITKKAHYIKSL 209

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 385

A DNA sequence (GBSx0416) was identified in *Sagalactiae* <SEQ ID 1253> which encodes the amino acid sequence <SEQ ID 1254>. This protein is predicted to be pyruvate dehydrogenase complex repressor. Analysis of this protein sequence reveals the following:

Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2827(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12044 GB:Z99105 similar to transcriptional regulator (GntR family) [Bacillus subtilis]
Identities = 67/225 (29%), Positives = 119/225 (52%), Gaps = 17/225 (7%)

Query: 3 RPLVEQTADRLGLILEREYVPGAKLPNEVEIARDLOVGRSTIREAVSLATRNLLVEVRQ 62
+ L +Q +R++HL+ + G KLP E EL + L V R +REA+ SL T ++ +
Sbjct: 16 ATLAKVIERIVHLISQQLRAGDKLPTEMLMDILHVSFVLEALREALSLETGLVITRAT 75

Query: 63 GSGTYISSKGVSEDPGFSLIKIDTDLTSDLPFLRLLEPRIAELWAYRITDQLOLLE 122
GTY + K G+ P L TD L + + E R+ LE + + A +I +++LQ L+
Sbjct: 76 RGSTYFNDKIGM--QPSVMLALATNLF-A-IIEARMAELGLVTTAEEKINEELRLQ 132

Query: 123 KMGDIKDAV--HAGDEKHLILLOVEFHSMLAKYSGNIAMDSLLPVIHQSHILNINANTYR 180
K + DI ++ H G+ D EFH +A+ A + N ++ ++ QS+ +A A ++
Sbjct: 133 KTIDDIANSTDNHYG----ADKEFFRIALISANNPVVVGMI----QSLITTHAKIDSQ 183

Query: 181 ---QMKSDSLEAHRELIKAIKEKNPVAAHDAMLMHMSVRRSALK 222
+ + ++E H++I A+ +++P AH N H+ VR LK
Sbjct: 184 IPIRERDVTVETHKIDYDALQDRPYKAHYHYELKFKVFDKILK 228

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1255> which encodes the amino acid sequence <SEQ ID 1256>. Analysis of this protein sequence reveals the following:

-489-

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

5 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2161 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

10 Identities = 24/51 (47%), Positives = 35/51 (68%)
 Query: 22 YPVGAKLPNEYELAEDELQVRSTIRKAVESLATRNILEVRQSGSTYISSKK 72
 +P+G++LP+E LAE V R T+R+A+ L ILE R GSGTY++S +
 15 Sbjct: 30 WPIGSRPLSERHIAEHFVSRMTLQAITLLVRGILERRIGSGTYVASHR 80

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 386

20 A DNA sequence (GBSx0417) was identified in *S. agalactiae* <SEQ ID 1257> which encodes the amino acid sequence <SEQ ID 1258>. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2178 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

30 A related GBS nucleic acid sequence <SEQ ID 9701> which encodes amino acid sequence <SEQ ID 9702> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA58911 GB:X84105 glucuronidase [synthetic construct]
 Identities = 258/602 (42%), Positives = 257/602 (58%), Gaps = 31/602 (5%)
 35 Query: 23 MLYPLLTKTNTYDLOGIWNFKLGEHNP-----ELLPSDEVMVITPSFNDAIWSKEK 75
 ML P+ T TR L G+W F L N L + +P SPND +
 Sbjct: 1 MLKPVTPTTRETIKKLDGLWAFSLDRNCGIDQRWWSALQBSRAIAPGSPNDQFADADI 60
 40 Query: 76 RDYIGDFWYKIEVPIKVSSEDEMVLKPGSVTHQAKIYVDGVLVGHEKGGFTFPEVLVPE 135
 R+Y G+ WY++ + +PK + +VLRF +VTH K++V+ V EH+GG+TPFE V
 Sbjct: 61 RNYAGNWNVQREVFIPKMGAGQRIVLRFDAVTHYGIKVVNNQVMEHQGGTYTFEADVTP 120
 45 Query: 136 CKYNEKIKVSI CANWFLDYTLVPVNYSLIIQDSGIKKVREMFDFNYAGVIRPKL 195
 + +++++C NN L+ T+P G I E+G KKK DFFNYG+HR + L
 Sbjct: 121 YVIAGKSVRTITVCVNNELNQITIPGMV--ITDENG--KKKQSYHDFNYAGIHRSVML 176
 50 Query: 196 MIRPKNHIFDITTSRLSDQLQSDALHFLAVETNKKVDEVRISVDFRDNKLV--GETKGR 253
 P + DIT+ + ++ D A + + V N +V+ + + D D + +V G+
 Sbjct: 177 YTTNTWVDDITVVTTHVQDCNHSVDWQVAVN--GDVSVELRADQVVVATQGTGT 233
 55 Query: 254 LPLSDVHLAEVLNAYLTARVETPVNQQLCVYENGLAREIVTNGQPLNARUPTVETG 313
 L + + HLA+ YLY V + D+Y G+R + V OPT+N KP YF G
 Sbjct: 234 LQVNNHMLQPGESYLYELCVTAQSQTCC-DIYPLRGIRSVAVKGEQLDNHKKPYFTG 292
 Query: 314 FQRHEDFPFGRLMEAAHMLADINLLKMGANSPRTSHYPTGSEEMRLADRGVLVIDEV 373
 FG+HED + G+G + + D L+ +GANS+RTSHYPT+KEM+ AD G++VIDE
 Sbjct: 293 FQRHEADLRGKGFNVLMVIDHALMDIGANSYRTSHYPTAEEMLLWADRGIVVIDET 352
 Query: 374 PAVGLFQNFNASLDS-----PKNGTWNLM--QTKAHHQATQELVKRDKRHPSVVM 425

-490-

AVG FN SL + PK+ + +T+ AH QAI+EL+ RDKNHPSVVMW
 SbJct: 353 AAVG----FNLSLGIGTPEAGNPKRELYSESAVNGETQGNHLQAIKELIARDKNHPSVVMW 408

Query: 426 VVNEPASHHEAGADYFELVLKYLKDLQKRFVTLVNLIMATPDRDQWMDVVLVCLNE 485
 +ANEP + GA +YF FL + + LDP RP+T VM++ D + DL DV+CLNR
 SbJct: 409 SIANSEFTRPQARSPAPLABATRKLDPT-RPTCVNVMPCDHYDTISDLFDVILCLNR 467

Query: 486 YYGWYVDHGLDINAEGVIRKELLWQDKFPDKPIIITEGADTLPLGLISTNYPIYTESFQ 545
 YYGWYV GDL AS + KELL WQ+K +PIIITEG DIL GLS + +HE+Q
 SbJct: 468 YYGWYVQSDLETPAKVLEKELLWQDKL-HQPIIITEGVDTLGLSHTYIMWSEHYQ 526

Query: 546 CIFYEMSHRVFDGIPNLVGRQVNFADPSTNIMILRVQGNHGLPSRNRQPKQVVKPEK 605
 C + +M HRVFD + +VGEQVNFADP T+ ILRV GN KG+P+R+R+PK +K
 SbJct: 527 CMLMDITHRVFDRVSAVVGQVNFADPATSQGLLRVGGNKGSIPTDKRKPASAEFLQR 586

Query: 606 RW 607
 RW
 SbJct: 587 RW 588

20 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1259> which encodes the amino acid sequence <SEQ ID 1260>. Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have no N-terminal signal sequence
 25 IYTBGRAL Likelihood = -4.04 Transmembrane 1131 -1147 (1130 -1147)

----- Final Results -----
 bacterial membrane --- Certainty=0.2614 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 30 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAF97242 GB:AF282987 beta-galactosidase precursor [Streptococcus pneumoniae]
 Identities = 303/921 (32%), Positives = 463/921 (49%), Gaps = 86/921 (9%)

Query: 5 QKSSRIVT---RTITKFSRATSNVKEIDMTFDSKBOTVTSQZYHYDQ--EGRKQPFH 58
 +K E VT + KP ++ + ++ E RK FH
 SbJct: 96 XKDEAVTPKEEKVSAKPEEKAPRIESQASNQEKFLKEDAKAVTNEEVNQMIEDRVDFH 155

Query: 59 GQWRF-LWADVACADPSPDSDNQVIHLPHDPSLTQPYTRNGEA--ESAYKLGQVGNRY 115
 Q W F L A+ A P D S W+ + LP+D+S+ + A E QG WYR
 SbJct: 156 QWYFKLANSGKRAIKPDADVSTWKLGLPYDMSIFNDPDHESPAQNBGGQLNGEAYNR 215

Query: 116 HYLVLDELVAGCHVAITFEGSYMETEITYVNGQPIGKHLNGYQRFYDTSIDVVF- GAENL 174
 LDE +V +TF+G YM+++YVNGQ+G + NGY +F+YDI+ + G EN+
 SbJct: 216 KTFKLDEKDLKNNVRLTFDGVYMDSGVYVNGQVGHYPYNGQPSYDITKYLQKDGRENV 275

Query: 175 LAHVVENKVPSSRWYSGSGLYREVSLSVLPLQHFVADQVAMTLADTAVQEGQQKVDYRF 234
 +AV NK PSRWYSGSG+YR+V+L V ++H + + O+ G+ + +
 SbJct: 276 IAVHAVNKQPSRWYSGSGIYRDVTLQVTKVHVKNGTTI+LTPKLERQHQGVETHVTS 335

Query: 235 ALNKGICTQHYQSLCINWQSHCSKDKKLLHYQTEFVLADAPQCYGLT--LSGETQL 292
 + + H ++ E + + L L + + +
 SbJct: 336 KIVNTDDKDELVA----EQYIVRGGHAVTGLVRTASRTLKASHTSLDALIVERPKI 391

Query: 293 WSP--DNEHLYDLELTLYQQQVIDCFCLTEQFQLTFMANQGLFVNGRAVLKGVCLHH 350
 W+ D P LY+L +Y QQ+D G+R + N+G +NG +K GV LHH
 SbJct: 392 WIVLNDKEALYELITRVYRQDQLVDAKKDLFGYRYHNTPNEGFSLNGERIKFHVGSYH 451

Query: 351 DQGLGACAYEDALARQLVLLKMGANTIRTSNPSPKLQLANRLGFPFVEEAFDIWT 410
 D G LGA A R+L +K+MG N+IR+THN+P+S + Q+L LG V EEARDFW
 SbJct: 452 DIGALGAENYKAEYRRLQKMGVNSIRTHIPASEQTLQIAELGLLVQESAFDIWT 511

Query: 411 YAKNGVNDPSNYPHOTIGTENANYLQVRVSPETSQWQYSIEMVWMAKNDPSVLMWSIG 470
 K D+ +F + A ++ W+ + + MV KN+P+ + MWSIG

-491-

	Sbjct: 512	GGK--KPYDYGRFFEKDATHPKARKGEK-----WSDFLRIMVERGKNNPAIFWWSIG	562
	Query: 471	NELMBGFSADVSHYELTRMQCQNTAIDTSRDTTFQDNKLKEADFC-WHREVSQAATLL	529
5		NE+ G + +H +++ + I +D +R +T G +K + + HE+++	
	Sbjct: 563	NEI--GEANGDAHSLATVKRLVKVIKVDKTRVVTMGADKFRPGNGSGCHKIA-----	614
	Query: 530	SQLDHPQGLIGLNYADGKDYDLHREHSDWLLYGSETVSAITSR-AYYKTKKVLDS---	585
		+LD +G NY++ +Y L +H WL+YGET SA +R +Y+ + + S	
10	Sbjct: 615	DELD---AVGFHYSE-DNYKALRAKHKWLLYGSETSSATRTGSSYRPERLKXNSGP	669
	Query: 586	--GYHL/SYDHAKVDMGAFASQAWYDTITTKDFV--AGRCVWTGFDYLGKPTPMWIKDSGV	641
		Y + Y + +V WG A+ +W T RD AG+ +WTG DY+GEPTW+ +	
15	Sbjct: 670	ERNYBQSDYGNDRVGWKTATASN--TFDRNAGYAQQFIWTGTDYIGEPTWNNQJPT	727
	Query: 642	VGLWPSPKNAYFGLIDTAGFEKDSYFFYQSQM--AQQSTLHLLFWQKQD-----QLCFD	694
		V K++YFGI+DTAG EK +Y YQSQM + + +HLLP W + +	
	Sbjct: 728	V-----KSSYFGIVDTAGIPKHDFLYQSQWVSVKKKMWHLLPHWNENKELASKVND	781
20	Query: 695	BQGLVEVVYNSNAASVQLMFEDEQGLIDYGRKAPHITYSTPTGHTYQLVQGDAAKNPHE	754
		+G + V YSNA+SV+L N G K F+ T G TYQ +GR+A	
	Sbjct: 782	SEGIKIPVARYSNASSVELFL---HGKSLGLKTPNKKQTSQGRITQ--EGANA-----	829
25	Query: 755	NLYLTWRVVPYQGLLRAVAYDISGKSIPTKSGRSQVTRYSVAKLSWKAPEAPIDAPW-E	813
		LYL W+V YQ G L A+D SGK I R + + T G A + + I A +	
	Sbjct: 830	ELYLEWVAVYQGTLELAJARDSSGKEI---ARDKITAGKPAVRLIKEDHAIAADQKD	885
	Query: 814	LLYLDLSLDERGELVSHAQMDLVQVGBPARLLALDNNPTDHTFYQEP-----LRQAY	868
		L Y+ +DS+G +V A + + Q+ G +L+ +DNG Y+ +R+A+	
30	Sbjct: 886	LTYIYYLVDQGNVFTANRLVRFLHGQQLGVGTNGEQASRERYKAQADGSWIRKAF	945
	Query: 869	GKLLALALTAGACHIKVTA 889	
		GK +AI+ T +AG +TA	
35	Sbjct: 946	NGKGAIVKSTEQACKFILT 966	

An alignment of the GAS and GBS proteins is shown below:

Identities = 98/414 (23%), Positives = 175/414 (41%), Gaps = 64/414 (15%)

	Query: 54	LPSDEVMVPIPTSPNDLMSKEKROYIDFWYKIEVKPVSEDERMVLRRGQSVTHAKYI	113
		LP D + P + N S K +G WY + + +V + + F + +IY	
40	Sbjct: 86	LPHDPSLTQPTPRNGARSAYKLGVG--WYKHYLVLDVELAGCHVAITPEGSTMETEYI	143
	Query: 114	VGVLVGEGKGGPTPFEVLVFECKYNNKKIKVSIKANVLYTILFVQNYSEIIRQDSI	173
		V+G +G+H G+ F + + V+ A N+L	
45	Sbjct: 144	VNGQFIGKHLNGYQETPTDSDV-----VTFGARILLAVR-----V	179
	Query: 174	KKKVRENPFDFNYAGVHRFLKLMIRPKNHIFDITITSRLSDLL-----QSADLHPLVET	227
		+KV + + + +G+R+ L + P+ H + L+D Q DL F +	
50	Sbjct: 180	ENKVPSS-RWYSGSGLYREVSLSVLQLHFLVADQVAMTLADTAVQEGQQVVDLRFALNQ	238
	Query: 228	NQKVDEVRISVF-----DEDNKLGRHTKDS-----RLFLSDVHLNVEIAT	267
		+ + + +S+ +D KL+ + + L L +LW N	
	Sbjct: 239	SIGTCHYQLS:CLMBQSHCSKOKKLLYQETVEPLADLAPRQYKGLISLEBELQWSPFNP	298
55	Query: 268	YLYTARVEIFDNDQLQVYENPGLREIE-VNMQPLIARKPIYFKGKHEDTFINGR	326
		+LY + + + Q+ D + G R++ + N +R + + RG H D G	
	Sbjct: 299	HLYDLSTLYYQQQGVDCPLETGFRLTPMANQGLFVNGRAVKI.KGVCLHHDQQLGRC	358
60	Query: 327	LNEAANI+DLNLLKMGANSFRTSHYFSEEMRLADRMQVLVIDEVPAGVLFQ---NPN	383
		E A L LKDMGN+ R+H P S + +L+R+G V+H+ + N N	
	Sbjct: 359	AYEDALANQLVLLKMGANTIRSTHNPSSPKLRQLANRGPFFVIREADFTWYAKNGVN	418
	Query: 384	ASLDLSPKDNQTN---LMQIAAH---BQIQELVKRDKNEPSVWVWVANE	430
		+ + + GF N L + + + +I+ +V KN PS+M+ +NE	
65	Sbjct: 419	DFSNYFCHTCTGTSNANYLQKRVSPETSNAQYSIEMVWSAKNDPSVLMWSIGS	472

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 387

A DNA sequence (GBSx0418) was identified in *S. galactiae* <SEQ ID 1261> which encodes the amino acid sequence <SEQ ID 1262>. This protein is predicted to be 2-keto-3-deoxygluconate kinase. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -0.53    Transmembrane    197 - 213 ( 197 - 213)

----- Final Results -----
      bacterial membrane --- Certainty=0.1213 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9699> which encodes amino acid sequence <SEQ ID 9700> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD35161 GB:AB001693 2-keto-3-deoxygluconate kinase (Thermotoga maritima)
Identities = 115/342 (33%), Positives = 180/342 (52%), Gaps = 16/342 (4%)

Query: 14 KIISLGEVILRLSPPOYHTLMQANHLLKQFGSGSEINVLASLAQLGYHVLVSALPDNDLG 73
      K+++ GE++LRLSPF + + Q + +GG+E NV A LAQ+G V+ LP+N LG
Sbjct: 2 KVVTFGRIMLRISPPDHKRIPTQDTSFDVYTGAGANVAAPLAQGLDLAYFTKLPRNPLG 61

Query: 74 KMASQFILSQQISPAAIKKRGLRIGIYYEYQGPSVRTNKVIYDRNYSPWSTLSYDFT 133
      A+ + + I + R+GIY+ E G S R +KV+YDR +S+ E+ D+D+
Sbjct: 62 DAAAGHLRIFGVKTDYIARQGNRIIGIYFLEIGASQRPSKIVYDRAHSAISEAKREDPDWE 121

Query: 134 SIFKGVDFWPHVSGITPALTKDLYEVTRFLMTKAKGGGVKVSIDLNPRESLNSSPQBARQ 193
      I G WPH SGITP L K+L + + A E GV VS DLN+R LW+ +EA++
Sbjct: 122 KILDGARWPHFSGITPPLGKELPLILEDALKVANEKGVTVSCDLNYARLWTK-REAQKV 180

Query: 194 LSPILGLLDVCFGLEPIYLAGESEDLAKDELGLSRPYLDI-----ELLEKITQKIVQRY 246
      + P + +DV L ED++ LG+S LD+ E KI +++ ++Y
Sbjct: 181 NIPFMEYVDV-----LIANEEDIEKVLGISVEGLDLTKLKNREAYAKIAEVRTKY 232

Query: 247 GLDYIAPFQREMEYTNQYMLKSYLYHNMLXQDKTGVEVLDKRVGTGDAPAGLHIALLE 306
      + T RE ++ N + +++ ++DRVG GD+FA LI+ L
Sbjct: 233 NFKTVGITLRESISATVMVSVMPFENGQPHFSNRYEIHIVDRVGAGDSFAGALITYGLIM 292

Query: 307 KETPQRALEIAMATFYKNIHTIQDINIMTRDDIAYLIEKSTN 348
      Q+ E A A KETI GD +++ ++I L T+
Sbjct: 293 GFDSQKKAEPMAASCLAHTIPGDFFVLSTIEETKLAAGATS 334
```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1263> which encodes the amino acid sequence <SEQ ID 1264>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0708 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 111/319 (34%), Positives = 168/319 (51%), Gaps = 7/319 (2%)
```

Query: 12 MAKIISLGEVLLRLSPFFQHTLMQANHLCQPGGSEINVLASLQGLYHVGLVSALEDND 71
 H+K++ +GE L+R+SP Q+ L A + FGGSE+H+ +H G L +ALFDN
 5 Subject: 14 MSKLLVGEPLIRVSPWQQLPLTNCEAQLFFGGSEVNIARTLGSGLEARLPTALFDNP 73

Query: 72 LGHVASQPTLSQQISPAAIKKKGRIGIYYEQQSPVTRNKVYIDRNVSEFWSTLSDDYD 131
 AG QP+ + + R+G+YY E QF R ++V YDR SSF D
 10 Subject: 74 VGHAFHQFLKQSGVDNLSLTAMQGHVGLVLYLENGFGCRASQVYIDRGGSSFSALDKDSD 133

Query: 132 FTSIPKGVDMFPVSGITPALTQDLRYVTRPLNFKAKBGQVKVSIDINPRESLMSSQEAR 191
 +IP+G+ FH SGI+ AL K ++ L+ +AK+ + +S DIAR G+ + +A+
 10 Subject: 134 LAAIFBGISHFPGISLAIKKTKTQDLIEVLVRAKGRDICSFDINPRESG-IAVADAK 192

Query: 192 EQLSPLI/LLDVCPGLRPVYLAGSEDLKDLGLSRFYLDIELEKITQKIVQRYGLDYI 251
 S D+ FG+EE+ L + D+ D R D + + + Q Y L I
 15 Subject: 193 RLPSPFAQYADIIFGMEPLLLSDDFDMFD-----RKADTTTIRKLAGLYQRYQLQAI 247

Query: 252 APTQREMEYTHQYMLKSYLHNHMLYQTDKTVGVVLRVGTGDAFAGLIHALLKSTPQ 311
 T+R + K+Y Y + Y++ + VL RVG+GDAF AGL+ LLE Q
 20 Subject: 248 YHTERNDAGGSHFKAYAY-DROPYSECVTTTFVLRVGGGDAFVAGLLYQLLEGNEQ 306

Query: 312 RALEIAGATPKYKHTIQGD 330
 R L+ A+AT K T+ D
 25 Subject: 307 RNLDFAVATSLKCTVAED 325

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 388

A DNA sequence (GBSx0419) was identified in *S. agalactiae* <SEQ ID 1265> which encodes the amino acid sequence <SEQ ID 1266>. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -1.17 Transmembrane 5 - 21 (5 - 21)
 35 ----- Final Results -----
 bacterial membrane --- Certainty=0.1468 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

40 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 389

45 A DNA sequence (GBSx0420) was identified in *S. agalactiae* <SEQ ID 1267> which encodes the amino acid sequence <SEQ ID 1268>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence
 50 INTEGRAL Likelihood = -12.05 Transmembrane 198 - 214 (191 - 220)
 INTEGRAL Likelihood = -11.68 Transmembrane 446 - 462 (437 - 467)
 INTEGRAL Likelihood = -9.55 Transmembrane 94 - 110 (91 - 116)
 INTEGRAL Likelihood = -7.43 Transmembrane 291 - 307 (283 - 309)
 INTEGRAL Likelihood = -4.88 Transmembrane 265 - 281 (257 - 282)
 INTEGRAL Likelihood = -4.62 Transmembrane 321 - 337 (318 - 339)
 55 INTEGRAL Likelihood = -3.93 Transmembrane 406 - 422 (405 - 426)
 INTEGRAL Likelihood = -1.59 Transmembrane 121 - 137 (121 - 137)

-494-

INTEGRAL Likelihood = -1.12 Transmembrane 345 - 361 (345 - 362)
 INTEGRAL Likelihood = -0.48 Transmembrane 43 - 59 (43 - 59)

----- Final Results -----

- 5 bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

- 10 >GP:CA13641 GB:Z99113 similar to H⁺-symporter [Bacillus subtilis]
 Identities = 105/452 (23%), Positives = 182/452 (40%), Gaps = 37/452 (8%)
- Query: 36 IYLETFMFVTFSTGVLSAAIFVSQIMGYIRIFDGFIDPAIGIMIDKTDTKFGKXRPIL 95
 IY ++ +F T V G +A + +RI D DP IG ++D+T+++F ++R L
 15 Sbjct: 27 IYATVSTYLLFFYTDFVGLSAAAAGTFLVVRILDALADPFGITIVDRINSRFARFPRYL 86
- Query: 96 IIGNVITALSLIFLLALRGVDENIRFFLFLVLIIHKIGYSMQQTITKQQTALINDPQ 155
 + G A + L L L + ++ I +G S+ T ALT+
 20 Sbjct: 87 LFG----AFFVLLAILCFTTDFPDSMEKLIYAYITTVGLSLTYTTINVPYGAITS-AMF 141
- Query: 156 RPIFNIVDAVMITSLMTGGQFVSVFLVKKGNFTPTQFFNVILFGTILISALIAIV--AI 213
 R +V L +V F VP + G L IL ++ +
 25 Sbjct: 142 RNNQEVSVISVRMLFANLGLVAVFVPLAALYLSDTSGNESLWQLTMGLILMGQCL 201
- Query: 214 IGIWAKDKREFFGLGENTQKTKALDKVWKLKGNKPLQILSIAAALVKFAIQFGFDSV-VN 272
 + K KE L ++ +K D ++ + N+PL +LSI ++ F + +SV +
 30 Sbjct: 202 LIPCFKSTKERVTLQKSEBKIKFTDIFEQFRVRLVPLVLSIFPIII--PGVNSIANSVOY 260
- Query: 273 VLLFGI----LFGVIALSGQFSLAFVGVIIINILPSTIARKKGLRFSYVRAIQIGMGL 328
 + + + L T L G L I+P L + +KK L + R+ + +IGL
 35 Sbjct: 261 VYTNLEREDLVKRWGLIGSLPALVILP--FPLRHQFLGKKLLNY----ALLMITGL 314
- Query: 329 LAPGAVLYGKPGDLSLTSINLYTILFVITNIIARYAQAPASLVLTGADISDYTES 388
 LA L + N+Y IL V +IA S + + + +Y +
 40 Sbjct: 315 LAL-----LFVPPSNVYLIL--VCRLIAAAGSLTAGGYMVALIPETIETGEXT 361
- Query: 389 GRIVSGMIGTIFSLTDSIASFPAMVVGFLAGIGFSKSPFTIETPLPDLMMAAISLV 448
 G+ + G+I I + +V G VL G+ P M +
 45 Sbjct: 362 GKRNGSLYIAIGFFPKFMALGVVPLVLDKFGY-----VANQAQTPAALM3ILITTT 416
- Query: 449 AIPFIALSIALLLMKFYKLDKEEMVRIGEKIQ 480
 IP L +AL+ + FY LD++ + + +
 Sbjct: 417 IIPVLLVLALIDINPYNLDEKYYIOMVRELE 448

- 45 No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 390

- 50 A DNA sequence (GBSx0422) was identified in *S. galactiae* <SEQ ID 1269> which encodes the amino acid sequence <SEQ ID 1270>. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 55 bacterial cytoplasm --- Certainty=0.3375(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

-495-

>GP:AAB17663 GB:U31175 D-specific D-2-hydroxyacid dehydrogenase [S. aureus]
 Identities = 165/331 (49%), Positives = 231/331 (68%), Gaps = 1/331 (0%)

Query: 1 MHKLKVENVREBZATLAQWANNHVELGMSGPIITLETVNEVBGPDQANAEIPLDDA 60
 M K+ F R+ E +A +W +H+VE++ S+ L+ TV++++ +DG+ Q L++
 Sbjct: 1 MTKIMFFGTDRYEKEMALWGKKNNVEVTTSSKELLSSATVDQLQYDGVTTWQFGKLEND 60

Query: 61 IYPLKEMGKIQIAQRGACVDWNIELAKHQGLIISNVPSYSPESIAEPTVTIATNLIK 120
 +YD L+ GIKQIAQR+AG DM+I+LAK+H I+ISNVPSYSPESIAE++V+TAL L+R+
 Sbjct: 61 VYFKLESYGKQLAQKTAGFDMYDLDLAKGNNIVISNVPSYSPETIAEYSVSIALQLVRR 120

Query: 123 VSLIRANVRBQIFSWLPLTRGRVLGNMFTVAILGTGRIGLATKIFKQPCRCVIGDYLYIN 180
 I V+ +F+W I + + NMIVAILGTGRIG ATAKI+ GPG + YD Y N
 Sbjct: 121 FPDIERVQAHDFTWQAEIMSKPVRNMIVAILGTGRIGATACTYAGPGATITAYDRYN 180

Query: 181 PMADGILETVNSVERAVERADIVSLHMPPTAEVTHLPNLWFKQFKGAILMMARGALV 240
 D L Y +SV+EA+++AD++SLH+P E+ HLF+ MF KKGAIL+N ARGA++
 Sbjct: 181 KDLD-FLTYKDSVKEAIKDADIISLRVFNKESYELFDKAMFDHVKKGAILVNAARGAVI 239

Query: 241 ETQDLLEALDQGLLEAGADITYEFGPYIPKNCQODISDSDPLRLNHKPYITTHAAY 300
 T DL+ A++ G L GA IDTYS E Y + -DI DK L LI H +++ +PH A+
 Sbjct: 240 NTFDLIAVNDGTLGAAIDTYNEAAVFTNDWINKDIDDKTLELIEHRIAVTPIHAF 299

Query: 301 YTCRAVQNLVEGALNACVEVIRTGTTITKVN 331
 ++CRAV+NLVEG LNA + VI TGT T++N
 Sbjct: 300 PSDEAVQNLVEGGLNALSINTGTCTRLN 330

There is also homology to SEQ ID 124.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 391

A DNA sequence (GBSx0423) was identified in *S.agalactiae* <SEQ ID 1271> which encodes the amino acid sequence <SEQ ID 1272>. Analysis of this protein sequence reveals the following:

Possible site: 40
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2364 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 392

A DNA sequence (GBSx0424) was identified in *S.agalactiae* <SEQ ID 1273> which encodes the amino acid sequence <SEQ ID 1274>. This protein is predicted to be regulatory protein (pfoS/R). Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood =-12.90 Transmembrane 64 - 80 (53 - 89)

----- Final Results -----

-496-

bacterial membrane --- Certainty=0.6158(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 5 A related GBS nucleic acid sequence <SEQ ID 9325> which encodes amino acid sequence <SEQ ID 9326> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AA65034 GB:AB001189 regulatory protein (pfoS/R) [Treponema
 pallidum]
 10 Identities = 33/91 (36%), Positives = 55/91 (60%), Gaps = 1/91 (1%)
 Query: 1 MANVLAKPKIMLFMISSAAILGILGALFNIQGTASAGFGISGLIGPINALNLAGKGSV 60
 M N + P + P++ + + G+L LFN+QGTASAGFG GL+GPINA L V
 Sbjct: 250 MPMWIRYPILNIPILLNGLVOGLANLFLNIGTASAGFGFGLGVOPINAYRLMAYTMMV 309
 15 Query: 61 MMMLIIIIIFVAAPILNIPFNFLFIKVLKI 91
 +L ++ FV + + ++ ++ + LK+
 Sbjct: 310 RAGILFLVYFVLS-FLAAYLIDFILVDRLKL 339

- 20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1275> which encodes the amino acid sequence <SEQ ID 1276>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have a cleavable N-term signal seq.
 25 INTEGRAL Likelihood = -12.31 Transmembrane 141 - 157 (133 - 166)
 INTEGRAL Likelihood = -6.00 Transmembrane 92 - 108 (88 - 112)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.5925(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 30 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AA65034 GB:AB001189 regulatory protein (pfoS/R) [Treponema
 pallidum]
 35 Identities = 63/178 (35%), Positives = 107/178 (59%), Gaps = 10/178 (5%)
 Query: 2 IQGGTASLLGLQPIIMSLIAMIFCPLIVSPITTVGIALAINLGGIGSGAASFG----- 55
 +G+ IA+ + LQP+IAS+L+M F +I+SP++V + +A+ L+G+ SGA+ G
 Sbjct: 164 VGRVIATPIALQPLIMSILLMSFSLIITISPVSSVAVGIAVGLGASGAANIGVSSCAM 223
 40 Query: 56 -LCLAGWAVNSKGTSLAHVLRSPKISMANVLSPKIMLFMLCSAAVLGVIGAIPIFNIGT 114
 L + VN G LA + K+ M N + P + P+L + V GV+ +EN+QGT
 Sbjct: 224 TLIVGTRVNMKIGVPLAFGAMMGLMFMWIRYPILNIPILLNGLVOGLANLFLNIGT 283
 45 Query: 115 ASAGPGISGLIGPINALNLAGKGSVCP-VNILLIIIIIFVGAPILNMIFNFLFIKVLKV 171
 ASAGFG GL+GPINA L + P V ++ ++ + + + ++ + LK+
 Sbjct: 284 ASAGFGFGLGVGINAYRIM--AYTMMVRAGILFLVYFVLSFLAAYLIDFILVDRLKL 339

An alignment of the GAS and GBS proteins is shown below:

50 Identities = 86/101 (85%), Positives = 96/101 (94%)
 Query: 1 MANVLAKPKIMLFMISSAAILGILGALFNIQGTASAGFGISGLIGPINALNLAGKGSV 60
 MANVL+KPKIMLFM+ SAA+LG++GA+FNIGTSPASAGFGISGLIGPINALNLAGKGSV
 55 Sbjct: 81 MANVLSPKIMLFMLCSAAVLGVIGAIPIFNIGTSPASAGFGISGLIGPINALNLAGKGSVCP 140
 Query: 61 MMMLIIIIIFVAAPILNIPFNFLFIKVLKVIDMDYKLDI 101
 +N+LLIIIIIFV API+AN LFNFLFIKVLK+IDMDYKLDI
 Sbjct: 141 VNILLIIIIIFVGAPILNMIFNFLFIKVLKVIDMDYKLDI 181

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 393

A DNA sequence (GBSx0426) was identified in *S. galactiae* <SEQ ID 1277> which encodes the amino acid sequence <SEQ ID 1278>. This protein is predicted to be regulatory protein (pfoS/R). Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -6.58	Transmembrane	148 - 164 (145 - 169)
INTEGRAL	Likelihood = -5.26	Transmembrane	33 - 49 (25 - 52)
INTEGRAL	Likelihood = -4.73	Transmembrane	70 - 86 (62 - 88)
INTEGRAL	Likelihood = -3.45	Transmembrane	124 - 140 (122 - 143)
INTEGRAL	Likelihood = -1.33	Transmembrane	96 - 112 (96 - 112)

----- Final Results -----

bacterial membrane	---	Certainty=0.3633 (Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

A related GBS amino acid sequence <SEQ ID 9735> which encodes amino acid sequence <SEQ ID 9736> was also identified.

A related GBS nucleic acid sequence <SEQ ID 9697> which encodes amino acid sequence <SEQ ID 9698> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AA065034 GB:AB001189 regulatory protein (pfoS/R) [Treponema pallidum]

Identities = 61/158 (38%), Positives = 92/158 (57%)

Query: 24 KSFIRNVINGLALQTVIVLIPGAILGELMALLFMWSGPATLIAATRVATSMGLVIGIM 83
+ F+M +LNG + G VI L+P AI GEL +AL P+ EA L + +IG +
Sbjct: 9 RQFMKILNGSSAGIVIGLVPPALAGELFALALPLSPLFALHVVLP IQFSVPALIGTL 68

Query: 84 VGLNPKFNPIQASGLAVMPAGCAATFLKQATLKGTDGINMGTITANLGVLLIQPLSD 143
VGL F + + A+L + A G T CA ++ G GD+IN+ + +AL ++L+ L
Sbjct: 69 VGLQPHCSAPEVATLAFVSVIASGNVTLQKAWLITGIGDIVNMWLISALAILVRLRG 128

Query: 144 KTKSPTLVIPTFTVLLNGGVGHVLLPVVMTITNIGQ 181
K S T+I +P + + GGVG LPVVMIT +G+
Sbjct: 129 KLSLTIIIALPVIVAVAGVGSPLSPVVMITLFPGR 166

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1279> which encodes the amino acid sequence <SEQ ID 1280>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -13.06	Transmembrane	314 - 330 (301 - 335)
INTEGRAL	Likelihood = -11.30	Transmembrane	185 - 201 (178 - 215)
INTEGRAL	Likelihood = -8.01	Transmembrane	22 - 38 (11 - 42)
INTEGRAL	Likelihood = -3.29	Transmembrane	266 - 282 (265 - 285)
INTEGRAL	Likelihood = -2.66	Transmembrane	141 - 157 (141 - 159)
INTEGRAL	Likelihood = -2.13	Transmembrane	53 - 69 (53 - 69)
INTEGRAL	Likelihood = -1.33	Transmembrane	114 - 130 (113 - 131)
INTEGRAL	Likelihood = -0.80	Transmembrane	206 - 222 (206 - 222)

----- Final Results -----

bacterial membrane	---	Certainty=0.6222 (Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>

-498-

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

5 >GP:AAC65034 GB:AB001189 regulatory protein (pfoS/R) [Treponema
pallidum]
Identities = 137/346 (39%), Positives = 217/346 (62%), Gaps = 14/346 (4%)

Query: 12 FMNKVLNGTAIAIVVALIPNAILATFLKPLLP-NMAAEFLHIVQVFQFTTPIMAGFLIG 70
FM K+L G++ IV+ L+P AI + L P + A H+V QF P + G L+G
10 Sbjct: 11 FMKTLNGSAGIVTGLVPPAIALGSLFRAIAPLSPLEALYHVVLPFQFSVPALIGTLVG 70

Query: 71 QQKPFNMQQLAVGGAAYIGSGAWAYTEVIQKGVATGTFQLRGIGDLINMITASLAVLA 130
QF + + + + I SG + G + + GIGD+IN+M+ ++LA++
15 Sbjct: 71 LQTHCSAPEVATLAFVSVIAGS-----NVTLQNGAWLITGIGDINVMILISALAIL 122

Query: 131 VKYFGNKFGSLTITLLPITITGTGVGYIGMKELPYVSYVTTLIGQGINSFTTLQFILMSIL 190
V+ K GSLTII LP+ + G +G LPYV +T +G+ I +F LQP+LMSIL
Sbjct: 123 VVALRGKLGSLTITLALPVTIVAVVGVSFSLPYVMITLFGVRVIALTIALQFILMSIL 182

Query: 191 IAVAFSLIIVSPISTVAIGLAIGLNGMAGAASMGIASTAAVLVWAILKNGKSVGPVIAIA 250
+++FSLII+SP+S+V+G+A+GL G+A+GAA++G++S A L+ T++VNK GVP+A+
20 Sbjct: 183 LMSFSLIITISPVSSVAVGIAVGLTGLASGAANTGVSSCAMTLIVGIMKVNKIGVPLAMP 242

Query: 251 LGAKMNMMPFLKHPIMAIPIWVFTAIISSLTIVPLFNLVGTASSGFLNGAVGVPAS--L 308
GAKM+MNP++++PI+ IP++ + + LFNL GTPAS+GFG +G VGP+ + L
25 Sbjct: 243 AGAKMMLNMPWIRYPILNIPLLNLGCGVLAWLNLGQTTPASAGFGFGLGVGPINAYRL 302

Query: 309 AGSSSIL---IIILAWIIVPFAFAAHKVKSDILKLYKEDIVFVE 351
++ ++ I+ L + ++ F A+ + D LKLY- ++F+ E
30 Sbjct: 303 NAYTFMVRAGILFLVIVFVLSFLAAYLIDFILVDRLKLYRRELFPI 348

An alignment of the GAS and GBS proteins is shown below:

Identities = 65/172 (37%), Positives = 95/172 (54%), Gaps = 9/172 (5%)

35 Query: 19 EKQTTSKPDINVLNGLALGTIVIVLIPGAILGELMALLPMMSGFATLIATAVATSNMGL 78
+K+T SF+ VL G R+ V+ LIP AIL +K ILP + A + V +
Sbjct: 5 DKETSSPMNKVLNGTAIAIVVALIPNAILATFLKPLLPNMAA-AEFLHIVQVFQFTTPI 63

Query: 79 VIGIMVGLNFKFNPISQASLGLAVMPAGGAATFLK-----GAINLGTGDIINMGIT 130
+ G ++G FKFNP+ ++G A GA + + G L+G GD+INM IT
40 Sbjct: 64 NAGFLIGQQKFNPMLQGLAVGGAAYIGSGAWAYTEVIQKGVATGTFQLRGIGDLINMIT 123

Query: 131 AALGVLLIQFLSDKTKSTFLIVIPFTVLLVGVGVHLLPYYVMITTMIGQ 182
A+L VL +++ +K S T+I++P VG +G LPYV +T+IGQ
45 Sbjct: 124 ASLAVLAVYFGNKFGSLTITLLPITITGTGVGYIGMKELPYVSYVTTLIGQ 175

A related GBS gene <SEQ ID 8567> and protein <SEQ ID 8568> were also identified. Analysis of this protein sequence reveals the following:

50 Lipop: Possible site: -1 Crend: 10
McG: Discrin Score: -13.49
GrH: Signal Score (-7.5): -5.82
Possible site: 48
>>> Seems to have no N-terminal signal sequence

55 ALOM program count: 5 value: -6.50 threshold: 0.0

INTEGRAL	Likelihood = -6.50	Transmembrane	148 - 164 (145 - 169)
INTEGRAL	Likelihood = -5.26	Transmembrane	33 - 49 (25 - 52)
INTEGRAL	Likelihood = -4.73	Transmembrane	70 - 86 (62 - 88)
INTEGRAL	Likelihood = -3.45	Transmembrane	124 - 140 (122 - 143)
INTEGRAL	Likelihood = -1.33	Transmembrane	96 - 112 (96 - 112)
PERIPHERAL	Likelihood = 1.85		51

60 modified ALOM score: 1.82

*** Reasoning Step: 3

----- Final Results -----

```

bacterial membrane --- Certainty=0.3633(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

5

The protein has homology with the following sequences in the databases:

OX601226 (352 - 843 of 1218)
 EGAD138195|TP00308|3 - 166 of 350) regulatory protein [Treponema pallidum] OXN1|TP0038
 regulatory protein (pfos/R) GP|3322295|gb|AAC65034.1|JAN001189 regulatory protein (pfos/R)
 [Treponema pallidum] PIR|E71373|E71373 probable regulatory protein (pfos/R) - syphilis
 spirochete
 %Match = 13.6
 %Identity = 37.2 %Similarity = 59.1
 Matches = 61 Mismatches = 67 Conservative Sub.s = 36

15

273 303 333 363 393 423 453 483
I•FFPFIPLQIAM•LI•LVKSQTIIISRRHLMEDVVEKQTTTSFIMNVLGALGTIVILIPAGAILGAELAKALLPWMS
 : : : : : : : : : : : : : :
MHTQSLSRPFWMKILGLSSSAGIVIGLVPPIAAGELFRALAPLSPL
 10 20 30 40

20

[illegible]

753 783 813 843 873 903 933 963
 SOKTKSFLILIVITPTVLITLVGGVGVHLLPVYVIMTTIGGGGRTHENFELILCFDINFEKIPF*INDLLSLFLQIIGL
 | | | : : : : : ||| ||||| : :
 RGLGLSLITIALPVIVAVVAGGVGSFLPVYVIMTILFVGVVIATFLQLPLLSILLMSFSLIISPVSSVAVGIAVL
 140 150 160 170 180 190 200

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 394

A DNA sequence (GBSx0428) was identified in *S. agalactiae* <SEQ ID 1281> which encodes the amino acid sequence <SEQ ID 1282>. This protein is predicted to be a cys operon transcriptional activator. Analysis of this protein sequence reveals the following:

40

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence
```

----- Final Results -----

45

```

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

50

>GP:CAE15857 GB:Z99123 alternate gene name: ipa-24d-similar to
transcriptional regulator (LysR family) [Bacillus subtilis]
Identities = 87/282 (30%), Positives = 152/282 (53%), Gaps = 5/282 (1%)

Query: 1 MDIRQLTYFIAVEAKNYSRAAKSLFVTQPTLSQSIKKLEAELNTILFLQNGRQLALTEA 60
MDIR LTYF+ VA K++++A++SL+V+QPT+S+ IK LE EL LF +NGRQ+ LT+A
Shift: 1 MDIRHLTYFLVARLKSPFKASQSLYVSQPTISKMIKNLBEELGIELFYRNGROVELTDA 60

22

Query: 61 GEILYKGGQLMTNVHQMVTIEIQQLNQEKKEGIRVGLTSLFAIQFMKQI-STFMATHSNV 119
G +Y + Q ++ + + +B+ ++ KK +R+GL + F ++ F + NV
Shift: 61 GHSMYVCAQEIISKPNLTSLNDIMEVKKGHVRIGLPPMIGSGFFPRVLGDFRENYENV 120

60

-500-

Query: 120 EVELIQDSRGLQGLANGIDIGLLSFPSTRNDITIEPLQTSTKGYKVSIVMPKSHPLA 179
 L++DGS K+OE + G +DIG++ P+ + + T + +V+ SH LA
 Sbjct: 121 TFLQVEDGSIKVBQSGVGGSLDIGVVLPAEDIPHSFTIVKET----LMLVHPHSRLA 176

5 Query: 180 TLPEIENLDRDYKVASLNHYMLGEMLPKRCRALGPDHPHIVFKHNDWEVLIHSLQDLNA 239
 E +L +L+D E ++L + +C GP PHI+++ +W+ + +
 Sbjct: 177 DEKCCQLRELKDEPFIFREDFVLHNRIMTICKAGFRPHIHYETSQNDPFISEMVSANLS 236

10 Query: 240 VTILPSFESISIQVDLCWFLAKDNNFYPIGIAVRNDTSFS 281
 + +LP + + +PL D + + I +R D S
 Sbjct: 237 IGLLPERICRGLDPEKVKVPLVDPVIFWHLAIWRKDRYLS 278

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1283> which encodes the amino acid sequence <SEQ ID 1284>. Analysis of this protein sequence reveals the following:

15 Possible site: 21
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 20 bacterial cytoplasm --- Certainty=0.1101 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 125/160 (78%), Positives = 144/160 (89%)

25 Query: 135 LAKGKIDIGLLSFPSTRNDITIEPLQTSTKGYKVSIVMPKSHPLATLPEIENLDRDYK 174
 L++GKIDIGLLSF S R DITIE LQTSTKGYKVSIV+ K HPLA P+++L DL+ YK+
 Sbjct: 1 LSQKIDIGLLSFLSIRKIDITIELLQTSTKGYKVSIVLLKQHPLAQHPLKLDLGYKI 60

30 Query: 195 ASLNHYMLGEMLPKRCRALGPDHPHIVFKHNDWEVLIHSLQDLNAVTLIPSEFESISQVQ 254
 ASLN+HYMLGEMLPKRCRALGP+P IVFKHNDWEVLIHSL DIAN +TILPS+PES+QV
 Sbjct: 61 ASLNHYMLGEMLPKRCRALGPEFDIVFKHNDWEVLIHSLDLNTLTILPSFESINQVD 120

35 Query: 255 DLCWFLAKDNNFYPIGIAVRNDTSFSFMIEEFLSLKTN 294
 +L W+PL+DKNNFYPIGIAVR+D SPSF+IEEFLSLKTN
 Sbjct: 121 NLWVIFLQDKNNFYPIGIAVRDASFSFVIEEFLSLKTN 160

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 395

A DNA sequence (GBSx0429) was identified in *S.agalactiae* <SEQ ID 1285> which encodes the amino acid sequence <SEQ ID 1286>. Analysis of this protein sequence reveals the following:

Possible site: 47
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1833 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

50 Signal peptide: 1-21

A related GBS nucleic acid sequence <SEQ ID 8569> which encodes amino acid sequence <SEQ ID 8570> was also identified.

55 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8570 (GBS271) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 8; MW 31.3kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 55 (lane 6; MW 56.3kDa) and in Figure 62 (lane 10; MW 56.3kDa).

- 5 GBS271-GST was purified as shown in Figure 210, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 396

- 10 A DNA sequence (GBSx0430) was identified in *S. agalactiae* <SEQ ID 1287> which encodes the amino acid sequence <SEQ ID 1288>. Analysis of this protein sequence reveals the following:

```

Possible site: 22
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -6.74 Transmembrane 9 - 25 ( 5 - 28)
INTEGRAL Likelihood = -5.84 Transmembrane 97 - 113 ( 92 - 122)
15 INTEGRAL Likelihood = -5.47 Transmembrane 37 - 53 ( 35 - 61)
INTEGRAL Likelihood = -2.55 Transmembrane 220 - 236 ( 220 - 238)
INTEGRAL Likelihood = -1.65 Transmembrane 64 - 80 ( 63 - 81)
INTEGRAL Likelihood = -1.28 Transmembrane 193 - 209 ( 192 - 209)
20 INTEGRAL Likelihood = -0.53 Transmembrane 125 - 141 ( 125 - 141)

----- Final Results -----
bacterial membrane --- Certainty=0.3697(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
25 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AA073593 GB:AE000155 putative metal resistance protein
[Escherichia coli K12]
Identities = 128/252 (50%), Positives = 186/252 (73%)

30 Query: 5 NSISLMSLMASSLVLTILFFSYQKLNLEKEVIGAIRAVIQLLAVGFLDYIFGYQNP 64
++I+ SL +A LV++ + S+ +KL LK++ S RA+QL+ VG++L YIF +
Sbjct: 13 HNITMSLALALMLVVAILSHKEKLALEKDIILNGVGRATIQILIVGVLYKTPSVDDA 72

35 Query: 65 IFTALMLFMINASYNAAKRGKINKQPVISFIAIGSTIITLSVLIPSGILKFPVPMQ 124
T L++L+P+ NA++NA KR K I K P+ SFIAI G ITH+VLI SG ++P+P Q+
Sbjct: 73 SLTLNVLVFCFNAWNNQKRSKYIAKAPISFIATTVGAGITLAVLISGSEIPMQV 132

40 Query: 125 IPVGMIISNSMVAIGLCYKQLSEFRSKQEEVFETKALGADILPASIDIIREDVKTGV 184
IP+ GMI N+MVA+GLCY L S+Q+++ KL+LGA AS +IRD I+ ++
Sbjct: 133 IPTAGMTAGNMMVAVGLCYNNLCQRVISBQQIQEKLGLGATPKQNSAILIRDSIRAAI 192

Query: 185 PTIDSAKTLGIVSLFGMMTGLLAGTEPIQAVKYQMMVTPLATTISIASFVATYLYKI 244
PT+DSAKT+G+VSLFGMM+GLI AG P++A+KYQ+MVTPLLT+T S++ +A YL Y+
45 Sbjct: 193 PTVDGAKTVGLVSLFGM+SGLI FAGIDPVKAIQYQMVTFMLSTISLTIACLYLYRK 252

Query: 245 FENNRRQLVVTK 256
F+N+R QLVT+
Sbjct: 253 FYNRRQLVVTQ 264
50

```

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 397

A DNA sequence (GBSx0431) was identified in *S.agalactiae* <SEQ ID 1289> which encodes the amino acid sequence <SEQ ID 1290>. This protein is predicted to be SUGAR TRANSPORT ATP-BINDING PROTEIN. (b0490). Analysis of this protein sequence reveals the following:

```

5      Possible site: 56
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.1903(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15      >GF:AACT73592 GB:AB000155 putative ATP-binding component of a
        transport system [Escherichia coli K12]
        Identities = 95/202 (47%), Positives = 142/202 (70%), Gaps = 2/202 (0%)

    Query: 4   LTFKHVDKIDDDKLVNDINFAIDBGFVSVIGPSGSGKSTVLKASGLMSPTAGHIFD 63
              L ++V + D +LN+INF++ G+F I GPSG GKST+LK+ + L+SPT+G + P+
20      Sbjct: 8   LQLQNGVLAGDAKILNNINFLRAGEFCLITGPSGCGKSTLLKIVASLISPTSGTLLEF 67

    Query: 64   GKDLNQLPEIESRNMISYCFQTHLFGNTVEDNISFPFYHIREHEKVDYRRVDDLFQRFEND 123
              G+D++ L+P R+ +SYC QTP LFG+TV DN+ FP+ IR+ + D D +RF +
25      Sbjct: 68   GBDVSLKPEIYRQVSYCAQTFILFGDTVYDNLIFPQIRNRQDPDAIFLDFLERFALP 127

    Query: 124  OSYLKQDVKKLSGGGKRIALIRQLLFEKPVLLLDVTSALDNHNKAIVEKVI -KSLHDK 182
              S L +++ +LGGGKQRI+LIR L F PKVLLDDE+TSALD NK V ++I + + ++
25      Sbjct: 128 DSILTKNLAELSGGKQRISSLIRNLQFMKPVLLLDDEITSALDESNNHNVNMIHYRVEQ 187

    Query: 183  GITILMITHDEEQSRFRANKVL 204
              I +LN+THD+++ A+KV+
30      Sbjct: 188 NIAVLAVTVHDKDEINM-ADKVI 208

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1291> which encodes the amino acid sequence <SEQ ID 1292>. Analysis of this protein sequence reveals the following:

```

35      Possible site: 61
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
40      bacterial cytoplasm --- Certainty=0.2053(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

45      Identities = 73/214 (34%), Positives = 133/214 (62%), Gaps = 9/214 (4%)

    Query: 4   LTFKHVD--FKTDDKLVNDINFAIDBGFVSVIGPSGSGKSTVLKASGLMSPTAGHIF 61
              +TF +V F+ Vb +INF ++BG F +++G SGSGKST+L + +GL+ ++G I+
50      Sbjct: 6   ITFNVSKTIFDSQTVLNRNINFDLEBGFYTLGLSGSGKSTILMIMAGLLDASSGDLY 65

    Query: 62   FDGKDLNQLPEIESRNMISYCFQTHLFGN -TVEDNISFPFYHIR--HKVDYRRVDDLPQ 118
              DG+ +N L PI R I PQ LF + TV +N++F ++ +K +RV + +
50      Sbjct: 66   LDGERINDL-PINKRD-IHTVPQNYALFHHMVFENVAFLKLLKVDKKEIAKRVKFTLK 123

    Query: 119  RFEMDSYLKQDVKKLSGGGKRIALIRQLLFEKPVLLLDVTSALDNHNKAIVEKVIK 178
              ++ + + + ++KLGGG+QR+A+ R ++ +P+V+LLE SAlD ++ + + +
55      Sbjct: 124 MVQL-BGFENRSIQKLGGGQRVAIARAIIHQFRVVLLEDEPLSALDLKATMEQYELR 182

    Query: 179  LHKD-GITILMITHDEEQSRFRANKVLKVNGBI 211
              L + GIT +++THD+E++ ++ + + G I
60      Sbjct: 183 LQRIIGITFFVFVHDEEALMSDNIFVNBGEI 216

```


Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 398

- 5 A DNA sequence (GBSx0432) was identified in *S.agalactiae* <SEQ ID 1293> which encodes the amino acid sequence <SEQ ID 1294>. Analysis of this protein sequence reveals the following:

Possible site: 32
>>> Seems to have no N-terminal signal sequence

- 10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.0658(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 399

- 20 A DNA sequence (GBSx0434) was identified in *S.agalactiae* <SEQ ID 1295> which encodes the amino acid sequence <SEQ ID 1296>. This protein is predicted to be dedA protein (dedA). Analysis of this protein sequence reveals the following:

Possible site: 58
>>> Seems to have no N-terminal signal sequence

- 25 INTEGRAL Likelihood = -12.05 Transmembrane 186 - 202 { 178 - 208 }
INTEGRAL Likelihood = -8.81 Transmembrane 65 - 81 { 61 - 89 }
INTEGRAL Likelihood = -7.54 Transmembrane 26 - 42 { 24 - 47 }
INTEGRAL Likelihood = -0.37 Transmembrane 152 - 168 { 152 - 168 }

- 30 ----- Final Results -----
bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 35 The protein has homology with the following sequences in the GENPEPT database:

>GP:AACT75377 GB:AB000320 orf, hypothetical protein [Escherichia coli K12]
Identities = 91/211 (43%), Positives = 131/211 (61%), Gaps = 7/211 (3%)

- 40 Query: 2 FLIDFLIHIDTHIYAMANTVGMWYLLPLVAVETGAVIFPFLPGDGLLFAAGLAANP 61
FLIDFLIHID H + + G W Y +LFL++P ETG V+ PFLPGDGLLFAAGLA+
Sbjct: 6 FLIDFLIHIDVHLAEIVAEYGVWVYAILFLFLPCTGLVVTVPFLPGDGLLFAAGLAASLE 65

- Query: 62 KMSFNIVTFLLIFPIAIPGDSNCFILGRITFGYRFIKHP---FFRRFIKEKHIRDAELYP 118
N+ +++ IAA+GD+ N+ IGR FG + +P FRR +K
45 Sbjct: 66 TNDLVNHHVVVLMILAAIVGDVAVYITGLPGEKLFSPNPKIPRSYLDK----THQFY 121

- Query: 119 EKKGTFAIILGRYIPILRTFVPFVAGISQLPPKVIKRAPIAALSNSVIATSGGFLGNI 178
EK G IIL R++PI+RTF PFAVAG + + F I AL W ++ T +G+ PG I
Sbjct: 122 EKKGGKTIILARFVPIVRTFAPFVAGGCHMSYRHPAAYNVIGALLVLLFTYAGYFPGTI 181

- 50 Query: 179 PFVKQHFSLILIGIVFVFLIPVLISGVKSYR 209
P V+ + L+I+GI+ V++P+ +I ++ R
Sbjct: 182 PMVQENLKILIVGIVVSLIPGVIRIRHR 212

-504-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 400

- 5 A DNA sequence (GBSx0435) was identified in *S.agalactiae* <SEQ ID 1297> which encodes the amino acid sequence <SEQ ID 1298>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3100(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

- 15 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 401

- 20 A DNA sequence (GBSx0436) was identified in *S.agalactiae* <SEQ ID 1299> which encodes the amino acid sequence <SEQ ID 1300>. This protein is predicted to be DNA-entry nuclease. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3990(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

- 30 A related GBS nucleic acid sequence <SEQ ID 9323> which encodes amino acid sequence <SEQ ID 9324> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

- ```
>GP:CAA38134 GB:X54225 membrane nuclease [Streptococcus pneumoniae]
Identities = 87/157 (55%), Positives = 110/157 (69%), Gaps = 1/157 (0%)

Query: 1 MLDRTIRQYQNRRTTLPDANKPLGWHQVAT-NDHYGHAVDKGHLIAYALAGNFKGWA 59
+L + RQY+NR++T +W P GWHQV Y HAVD+GHL+ YAL G G+DA
Sbjct: 116 LLSKATRQYQNRKKTGSGTSWTFPGWHQVKNLKGSTHAVDRGHLLGVALIGSLGDFDA 175

Query: 60 SVSNPQNVVTQTALSHNSQNKINRGONYYESLVKRAVDONKRVRYRVITPLYRNDTLVPF 119
S SNP+N+ QEA +NQ+ + + GQNYYES VRKA+DONKRVRYRVIT Y ++ DLVP
Sbjct: 176 STSNPKNIANVTAWANQAQAEYSTGQNYYESKVRKALDONKRVRYRVITLYASNEDLVPS 235

Query: 120 ASGLEAKSGDGTLEFNVVAIPHTQASTYMDYATGEITL 156
A +SAKS DG LEFNV +PN Q +DY TGE+T+
Sbjct: 236 ASQIEAKSSDGLLEFNVVAVNQKGLQDYRTGEVTV 272
```

- 50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1301> which encodes the amino acid sequence <SEQ ID 1302>. Analysis of this protein sequence reveals the following:

-505-

Possible site: 42  
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----  
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAA38134 GB:X54225 membrane nuclease [Streptococcus pneumoniae]  
 Identities = 89/135 (65%), Positives = 104/135 (76%), Gaps = 1/135 (0%)  
 Query: 25 SPAGWEHLHLKGSYTHAVDRGHLLGYALVGLGFDASTGNPDNIATQLSWANQANKPY 84  
 +P QNH++ +LKGSY HAVDRGHLLGYAL+GGL GFDAST NP NIA Q +WANQA Y  
 Sbjet: 138 TTPGWHLQVKNLKGSYTHAVDRGHLLGYALIGLGDFASTSNPKNIATQTMANQAQNEY 197  
 Query: 85 LTQGNYYEGLVRRALDKGHRVRYRVTLTY-DGNLLASGSHLEAKSSDDSLTFNVFVNV 143  
 TQGNYYE VR+ALD+ RVRVRYVTL Y ++L+ S S +EAKSSD L FNV VENV  
 Sbjet: 198 STQGNYYESKVRKALDKNKRVRVRYVTLTYASNDLVPSASQIEAKSSDGLFNVFVNV 257  
 Query: 144 QAGLTADYRTGQIAI 158  
 Q GL DYRTG++ +  
 Sbjet: 258 QKGLQLDYRTGQIVTV 272

An alignment of the GAS and GBS proteins is shown below:

Identities = 73/135 (54%), Positives = 92/135 (68%), Gaps = 2/135 (1%)  
 Query: 24 PLGNHQA-INDHYGHAVDKGHLIAYLAGNFKGNDASVNPQNVVTCTAHSNQSQKIN 82  
 P QNH++ Y HAVD+GHL+ YAL G KG+DAS NP N+ TQ + +NQ+N+  
 Sbjet: 26 PAGNHLHLHLKGSYTHAVDRGHLLGYALVGLGFDASTGNPDNIATQLSWANQANKPYL 85  
 Query: 83 RGQNYYESLVKKAVDQNKRVRYRVTPLYRNDITLVFPFAMHLEAKSQDGLTFNVFVNV 142  
 GQNYE LVR+A+D+ RVRVRYVTL Y D +L+ HLEAKS D +L FNV +FN Q  
 Sbjet: 86 TGQNYEGLVRRALDKGHRVRYRVTLTYDGD-NLLASGSHLEAKSSDDSLTFNVFVNV 144  
 Query: 143 ASYIMDYATGSEITLN 157  
 A T DY TG+I +N  
 Sbjet: 145 AGLTADYRTGQIAIN 159

SEQ ID 9324 (GBS656) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 186 (lane 10; MW 57kDa).

GBS656-GST was purified as shown in Figure 236, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 402

A DNA sequence (GBSx0437) was identified in *S.galactiae* <SEQ ID 1303> which encodes the amino acid sequence <SEQ ID 1304>. Analysis of this protein sequence reveals the following:

Possible site: 13  
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----  
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

-506-

A related GBS nucleic acid sequence <SEQ ID 9321> which encodes amino acid sequence <SEQ ID 9322> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1305> which encodes the amino acid sequence <SEQ ID 1306>. Analysis of this protein sequence reveals the following:

```

Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.5350(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 24/73 (32%), Positives = 37/73 (49%), Gaps = 2/73 (2%)

Query: 1 MFYMKIANRLSLAATIVNKAANSPPGLIIHSDKARNVWINDFETQFPDLNSPKKESSP 60
 + YMKLA L TI+ E + SFF I+H+D A N++ E N +++P
Sbjct: 80 ILYMKIAKRNHLFVITITETHMTSPFAFILYTHRAINLKETRLKLVILKQTINDQLSKQTF 139

Query: 61 K--KSLNQHFPPSQ 71
 + KS N+ F +
Sbjct: 140 ERTKSPWKRFLKK 152

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 403

A DNA sequence (GBSx0438) was identified in *S.galactiae* <SEQ ID 1307> which encodes the amino acid sequence <SEQ ID 1308>. This protein is predicted to be Isopentenyl-diphosphate delta-isomerase.

Analysis of this protein sequence reveals the following:

```

Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1649(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AFG20030 GB:AE005083 isopentenyl pyrophosphate isomerase; Idi
[Halobacterium sp. NRC-1]
Identities = 24/77 (31%), Positives = 40/77 (51%)

Query: 14 TGLTLNRDQNIPOSLFHLVVDVILFHEDGDVLNMRKHEKKAFAYFEATAGGSALKGEN 73
 TGL D + G+ H +L+ EDG VL+ +R +K+ + +++ T ++G++
Sbjct: 42 TGLANRLDAHTGDGVHRAFTCLLFDIEDGRVLLAQRADRKRLNDTHWDGTVASHPIBGQS 101

Query: 74 AKQAILRELKEETGIVE 90
 A + L EE GI F
Sbjct: 102 QVDATKRLABELGIEP 118

```

No corresponding DNA sequence was identified in *S.pyogenes*.

-507-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 404

A DNA sequence (GBSx0439) was identified in *S.agalactiae* <SEQ ID 1309> which encodes the amino acid sequence <SEQ ID 1310>. This protein is predicted to be phosphoserine phosphatase (serB). Analysis of this protein sequence reveals the following:

Possible site: 35  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0613 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB50876 GB:AL096844 putative phosphoserine phosphatase  
[Streptomyces coelicolor A3(2)]  
Identities = 96/193 (49%), Positives = 132/193 (67%)

Query: 5 LLVMDVDSSTLIMBEAIDLLAIEAGVGKQVAALTYAAMRSELDPEEALIKKRVALLKGLFVT 64  
L+VMDVDSSTLI +E I+L A AG +VA +T AAMRSELDPE++L RVALL GL +  
Sbjct: 183 LVVMDVDSSTLIQDEVIELFAAHAGCDEVAEVTAAAMRSELDPEQSLHARVALLAGLDAS 242

Query: 65 ILTDIISIIHPTFGAYELIKECHKRQMKVGLVSGGFHETIDILAKQLQVDVYKANRLGVK 124  
++ + + + TPGA LI+ + +VG+VSGGF + D L +QL +D+ +AN L +  
Sbjct: 243 VVDKVRAEVKITPGARTLIRTKRLGYQGVVSGGPTQVTDALQEQGLDPAQANTLEIV 302

Query: 125 GSPFTGVRGSEIVTKVEVKIKLKEWASENHLDLSQTAMGDGANDLPMIKSAGGVIAFCA 184  
G LTG V GETV + K L+ +A+ +LSQT+A+GDGANDL M+ +AG+G+AF A  
Sbjct: 303 DGRITGRVIGETVDRAGKARLRRFAAAGVPLSQTVAIGDGANDLMIANANGLVAFNA 362

Query: 185 KPIVREAAAYQIN 197  
KP+VRE A +N  
Sbjct: 363 KPVRREAARTAVN 375

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 405

A DNA sequence (GBSx0440) was identified in *S.agalactiae* <SEQ ID 1311> which encodes the amino acid sequence <SEQ ID 1312>. Analysis of this protein sequence reveals the following:

Possible site: 23  
>>> Seems to have an uncleavable N-term signal seq  
INTERGRAL Likelihood =-17.88 Transmembrane 5 - 21 ( 1 - 29)

----- Final Results -----

bacterial membrane --- Certainty=0.8153 (Affirmative) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB06924 GB:AP001518 unknown conserved protein [Bacillus halodurans]  
Identities = 122/553 (22%), Positives = 265/553 (47%), Gaps = 12/553 (2%)

-508-

Query: 7 LLLVAIVLLVIAIYVGVIRKNDITIANLESTRQQLVDPVQSEIBQVKLLHLIGQSQ 66  
 +++ ++++L + +V G + RK + LE K +++ P+ +EI +VK L + G+++  
 Sbjct: 3 IIVFSLVLTIVTFVFGALRRKAFYKRVKLEWKNILQRPIDPEIGKVGKLTMSGETE 62

5 Query: 67 STFRWNQKQWDLSTNSFKDIDPHLVKRNINDSFNFVRAKHSDINVSQLTIERHDIVS 126  
 F W W D+ +++ L + E+ + + F +AK +D ++ +L IEE +  
 Sbjct: 63 EKFEVWRSDWDIDVGVILPNVBEQLFDVSDFANKYRPFQKAKALLDTIEQLHSIREQLKI 122

10 Query: 127 IREAIHVLEKQEKONSARVTHALDLYETLQKSISEKNDNYGTTMPSEKQLANIEAFESH 186  
 + + ++VL + EE+N + +L + L K + + + ++ ++L+  
 Sbjct: 123 MVDDIQVLQSEBQNRTEIGSVRELQOKLIKRAITRSGSLSSAKVPDEKLEKANELLQA 182

Query: 187 FVFLINSTQDPIASRVIMKABEHTIALQQTIEQIPAIKALDDPPQQLDLETGYRRL 246  
 F G+ I+ASEVL +A+E + + + +P + +L+ + P +L +L+ G R +  
 Sbjct: 183 FDEKTEKGNITQASEVLEAKELIQIEHLIKIVPGLFVQLQTNIPAKLNLKNGLRDNE 242

15 Query: 247 EBNYHFFKEDIBQRFQVREAIRSNSDGLVSLDLDRADENSHIQSKIDCLYDIEREIA 306  
 E + + I+ + + + E + L L+ + +E I+E ++++++ E+E+  
 Sbjct: 243 EAGFFLETPAIDSQMERLEEKRVELLEQLTVLEKNGMEESINFTEESBQMPFLEKEVE 302

20 Query: 307 AYKVAHKCSKIIPQFLAHAKSNNEQCH---EIKFLSAKYILNENESLSRSFTNDLEEI 363  
 A + + + +P E+L H E + + Y L E E + + +L+E+  
 Sbjct: 303 A---KNEITILLPNLRDLTKTEKLLHKEETESVQLSYRLAEELVPQKLGKEL 359

25 Query: 364 ETKVLSVENPQZASPTYLQILPFTLTKTLTVENKQMEVFRAVTKTIESVTRARQM 423  
 ++ E ++ ++ ++ + E + LT + + E ++ + E + ++ +  
 Sbjct: 360 RQQLQVIDEVTEEQKQTFSSVRNLEBWRGLTAQKNIBQQRSLNLSLRKDKAKEEL 419

30 Query: 424 DKYVNLKLMKRPMEKRNLEPGIPQDPLSTFTTSSQIALINELSGRGRIDIAVSRNDV 483  
 + KL KR ++K N+PG+P+ L + + I +LS +++ V+ L D  
 Sbjct: 420 KQLKEKLEDKRLVQKSNIPGLPETLHRLDEQKLAIAKLSDVPLEMRVLTALVDE 479

Query: 484 TTNALMLEQATYLVQDAILTEQLQSNRYSPFQVKNQFQALVLFVEVHNKASF 543  
 I + + ++ A L E ++QV NRYRS V+K A L F +  
 Sbjct: 480 AQGLIHENSILHETIEKARLAEHVITQGNRYRSRAEVKRLNVAELFRA-----FEY 534

35 Query: 544 DE-ISALETVEP 555  
 DE I A++ +KP  
 Sbjct: 535 DEAIEMAVQAIEP 547

40

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1313> which encodes the amino acid sequence <SEQ ID 1314>. Analysis of this protein sequence reveals the following:

Possible site: 23  
 >>> Seems to have an uncleavable N-term signal seq  
 45 INTEGRAL Likelihood = -18.04 Transmembrane 5 - 21 ( 1 - 29)

----- Final Results -----  
 bacterial membrane --- Certainty=0.8217 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 50 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GB:BA06924 GB:AP001518 unknown conserved protein [Bacillus helodurans]  
 Identities = 131/555 (23%), Positives = 269/555 (47%), Gaps = 16/555 (2%)

55 Query: 7 LLLVAIVLLVIAIYVGVIRKNDISLTSLEKQQLFALVNDIEBVKSLHLIGQSQ 66  
 +++ ++++L + ++ G + RK + LE+ K + P+ DEI +VK L + G+++  
 Sbjct: 3 IIVFSLVLTIVTFVFGALRRKAFYKRVKLEWKNILQRPIDPEIGKVGKLTMSGETE 62

60 Query: 67 STFRWNQKQWDLTVNSFADIEKNIPRAENLNDTFNIRAKHESINVSQNLNVEEDIAS 126  
 F W W D+ ++E +P+ E+ + + F +AK ++++E +L+ +E +  
 Sbjct: 63 EKFEVWRSDWDIDVGVILPNVBEQLFDVSDFANKYRPFQKAKALLDTIEQLHSIREQLKI 122

Query: 127 IREAILNLEKQEKONSARVTHALDLYEKLQASISEKNDNPGSTMPRIKQMONITEFSQ 186  
 + + + +L + EE+N + +L +KL + S+ D++++  
 65

Sbjct: 123 MVDIDQLVQSEQNTTEIGSVRELQQLIKRAITRRGSLSSAKVFDEKLEKANELLQA 182

Query: 187 FVALNSSGDPVASEVLDRAREHTIALGQITEQIPAIKLEDDFPQQLDLETGYRRLL 246  
 F G+ ++ASEVL+ A+E + + + +P+ +L+ +P +L+ G R +

5 Sbjct: 183 FDERTEKGNVQASEVLEBAKELLAQIEHLKIVPGLFVSLQNTI PAELINLADMB 242

Query: 247 EENYHFFPKNIARFPQEIRESIRANSSELVTLDDRARENTHIQERIDSLYEVFERIA 306  
 E + + 1+++ + + E +L L+ + EE I+E + + ++E+ E+E

10 Sbjct: 243 EAGFFLETVAIDSQMERLEKKRVLELQITVLECNMEEELNFTESNEMFPELEKE-- 300

Query: 307 AYKVAKN--SKMLPYLEHVKRNNEQ---LKDEIARLSRKYLSSTESLVKFAEKDIK 361  
 V AKN + +LP E + + E+ LK+E + Y L+E E + + K+K

Sbjct: 301 ---VEAKNETITLPAULREDLTKTEKLTHLKRETESVQLSYRLAEELVFPQKLGEKAK 357

Query: 362 EIEDSTLAVABQFGLQEKPFSELQVTTFERSIKTLNIVESQMDVFAVVDIEKIESQARH 421  
 E+ + E Q++ FS ++ E + LE ++ ++ + K E +A+

15 Sbjct: 358 ELAQQLQVIDETEEQKQTFSSVRSHLEEWREGLTAQCNKIBQQSLSLRKDELKAKE 417

Query: 422 NLDVYVTLQHMIKRYMEKRLPGIPQDFLSAFFITSSQLEALDELSGRINIEAVSRLS 481  
 L +L KR ++K +PG+P+ L +L + +LS + + +V L

20 Sbjct: 418 ELKQLKEKLEDKRLVQKSNIPGLPETLHRLDEGEQKLAQAIKLSVPLEMGRVTLV 477

Query: 482 EVATVAIANLEDLTYVQVATLTTEQLQYSNRYSPFQVQSSFEHALFVENDYQA 541  
 + A I + + + + A L E + +QY NRYES A V+ +A LP

25 Sbjct: 478 DEAQQLIHENSSILHETIEKARLAEHVQYGNRYSRSAEVKKRLNABELFRA-----F 532

Query: 542 SFDR-ISALETVEP 555  
 +DE I A++ +EP

30 Sbjct: 533 EYDRAIEMAVQAIEP 547

An alignment of the GAS and GBS proteins is shown below:

Identities = 429/574 (74%), Positives = 503/574 (86%)

Query: 1 MSSGIILLVAIVLLVVIAYVGVVIRKRNQTLIANLETRKQELVOLPQEQERIQVKILH 60  
 MSSGIILLVAIVLLVVIAYVGV+IRKRN+DI +LE RKQ L L FV +EIB+VK LH

35 Sbjct: 1 MSSGIILLVAIVLLVVIAYLVGVVIRKRNDSLTSLSEKQALFALPNDIEIEVSKSH 60

Query: 61 LIQSQQTSFREDNQWVLTSTNSFKDIFHLVEARNLDSFNFVRAKHEIDNVDSQLTI 120  
 LIQSQQ++FREDNQW DL+ NSP DI+ H+ EAENLD+FN+RAKHEI++V+SQL ++

40 Sbjct: 61 LIQSQQTSFREDNQWVLTSTNSFADIENHIFAEENLDNTPNFIRAKHEINSVESQIALV 120

Query: 121 EEDIVSIREALEVLKQEQEKNARSVTHALDLYETKQISSEKEDNQTMTPEIRQLKNI 180  
 EDI SIREAL +LWQEQEKNARSVTHALDLYE LQ SISE EDN+G+TMPEI+Q+RN

45 Sbjct: 121 EEDIASTIREALTKRQEQEKNARSVTHALDLYETKQASISENEDNGSTNPEIDQMKNI 180

Query: 181 EAEPSHPTVLTSTGDIPEASEVLMAKAEHTIALGQITEQIPAIKLEDDFPQQLDLET 240  
 E EPS FV LNS+GDP+EASEVL++AEHTIALGQITEQIPAIKLEDDFPQQLDLET

60 Sbjct: 181 ETEFSPQVVALNSSGDPVASEVLDRAREHTIALGQITEQIPAIKLEDDFPQQLDLET 240

Query: 241 GYRRLEENYHFFPKDIQRPOFQVREAIRNSDGLVSLDLDRARDNEHIQEKIDKLYDI 300  
 GYRRLEENYHFFPK+IE RPO+RE+IR+NS LV+LDDRAR+EN HIQ+D LY++

50 Sbjct: 241 GYRRLEENYHFFPKNIARFPQEIRESIRANSSELVTLDDRARENTHIQERIDSLYEV 300

Query: 301 PEREIAAYKVAHKDSKIIPOPLWAKENNEBQCHEIKRLKQYHIANENESLSLSPNDL 360  
 PEREIAAYKVA K+S++P++L H K NNEQL EI RLS KYL+E ESL++++P D+

55 Sbjct: 301 PEREIAAYKVAAKNSKMLPYLEHVKRNNEBQKDEIARLSRKYLIESTESLVKFAEKDI 360

Query: 361 KEIEIKVLPVSNPQGEASPYTHLQILFERTLKLITVTEENQVFEAVKTESVETRAR 420  
 +EIE L E PG + P++ LQ+ FRK++XTLT VS QN+VF AVK IE +E+AR

60 Sbjct: 361 KEIEDSTLAVABQFGLQEKPFSELQVTTFERSIKTLNIVESQMDVFAVVDIEKIESQAR 420

Query: 421 QNMDKYVNLKIMIKRPMKRNLPQIPQDFLSTPFTTSQIBALINELSGRIDIEAVSR 480  
 N+D YV +LHMIR+MKKR+LPGIPQDFLS FTTTSSQ+KAL++KLSGRINIEAVSR

65 Sbjct: 421 HNLDVYVTLQHMIKRYMEKRLPGIPQDFLSAFFITSSQLEALDELSGRINIEAVSR 480

Query: 481 NDVTVAIANLEBATYLVQDATLTTEQLQYSNRYSPFQVQSSFEQALYLFVEHNYK 540  
 ++V T AIANLE TY VVQ+ATLTTEQLQYSNRYSPFE VQ SE AL LFEVE++Y+

-510-

Sbjct: 481 SEVATVAIANLEDLTYQVQVQATLLEQLLQSNRYRSFZAGVQSSPEHALRLPVEVDYQ 540

Query: 541 ASFDEISYALETVEPGVTRFVTSYKTKRRIRF 574

ASFDEISYALETVEPGVTRFV SYET+R IRF

Sbjct: 541 ASFDEISYALETVEPGVTRFVNSYKTRSHIRF 574

SEQ ID 1312 (GBS642) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 142 (lane 2-4; MW 27kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 406

A DNA sequence (GBSx0441) was identified in *S. agalactiae* <SEQ ID 1315> which encodes the amino acid sequence <SEQ ID 1316>. Analysis of this protein sequence reveals the following:

Possible site: 41

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2471 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9671> which encodes amino acid sequence <SEQ ID 9672> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>CPA91553 GB:E67740 DNA gyrase [Streptococcus pneumoniae]

Identities = 574/650 (88%), Positives = 618/650 (94%), Gaps = 2/650 (0%)

Query: 1 MTEETKNMSQRAQYDASQIQVLEGLRAVMRPGMYIGSTSKBGLHHVWEIVDINSIDEA 60

MTEE KN++ AQ+YDASQIQVLEGLRAVMRPGMYIGSTSKBGLHHVWEIVDINSIDEA

Sbjct: 1 MTESEIKNLQ--AQDYDASQIQVLEGLRAVMRPGMYIGSTSKBGLHHVWEIVDINSIDEA 58

Query: 61 LAGFAGHIKVYIEPNSITVDDGRGIPVDIQEKTGRFAVETVFTVLHAGGKFGGGGYKV 120

LAGFA HI+V+IEPD+SITVDDGRGIPVDIQEKTGRFAVETVFTVLHAGGKFGGGGYKV

Sbjct: 59 LAGFASHIQVIEPDSITVDDGRGIPVDIQEKTGRFAVETVFTVLHAGGKFGGGGYKV 118

Query: 121 SGGLHGVSSVNALSTQLDVKVYKNGKVHYOEYQRGVVNDLEIIGDITLGGTVRFTF 180

SGGLHGVSSVNALSTQLD V+KNGK+HYOEY+RG VV DLE++GDTD +GTTVRFTF

Sbjct: 119 SGGLHGVSSVNALSTQLD VHVHKNKGIHYEYRGRHVADLEVVGDTDTGTGTVRFTF 178

Query: 181 DPEIFTETTVPDFDKLAKRIQELAFINRGLRISIDSKRGQVKEKSYHYEGGIGSYVEFI 240

DPEIFTETT+PDFDKL KRIQELAFINRGL+ISI+DKR+G E K YHYEGGI SYVE+I

Sbjct: 179 DPEIFTETTVPDFDKLNKRIQELAFINRGLQISITDKRGLQETKHYYHGGIASYVEFI 238

Query: 241 NENKEVIFENPIYTDGELDGISVEVAMQYTTGYQETVMSFANNIHTHEGGTHEQGFRTAL 300

NENK+VIP+ PIYTDGE+D I+VEVAMQYTTGY E VMSFANNIHTHEGGTHEQGFRTAL

Sbjct: 239 NENKDVIFUTPIYTDGEMDITVEVAMQYTTGYHENVMSFANNIHTHEGGTHEQGFRTAL 298

Query: 301 TRVINDYAKGKILKGNEDNLTGEDVREGLTAVISVKHPNQFBSQIKTKLGNSEVVKIT 360

TRVINDYA+KHK+LK+NEGNLTGEDVREGLTAVISVKHPNQFBSQIKTKLGNSEVVKIT

Sbjct: 299 TRVINDYAKGKILKGNEDNLTGEDVREGLTAVISVKHPNQFBSQIKTKLGNSEVVKIT 358

Query: 361 NRLFSEAFNRFLENEQVAKIVKGIILASKARIAAKRAREVTRKKSGLIENLPGLKLD 420

NRLFSEAF+ FL+KNEQ+AK+IVKGIILA+KAR+AAKRAREVTRKKSGLIENLPGLKLD

Sbjct: 359 NRLFSEAFSDFLKNEQIAKRIKVEKGIILAAKARVAAKRAREVTRKKSGLIENLPGLKLD 418

Query: 421 CSNNNAEMNELFIVEGDSAGGSAKSGRNREFQAILPIRKILNVEKATMDKILANERIS 480

CSNNN RLFIIVRGDSAGGSAKSGRNREFQAILPIRKILNVEKATMDKILANERIS

Sbjct: 419 CSNNNPASTELFIVEGDSAGGSAKSGRNREFQAILPIRKILNVEKATMDKILANERIS 478



-511-

Query: 481 LFTAMGTGFGADFDVSKVRYQKLVIMTDADVDGAHIRTLLLTLLIYRMPRVLEAGYVYIA 540  
 LFTAMGTGFGA+FDVSK RYQKLV+MTADVDGAHIRTLLLTLLIYR+M+P+LEAGYVYIA  
 Sbjct: 479 LFTAMGTGFGARFDVSKARTYQKLVIMTDADVDGAHIRTLLLTLLIYRINKPYLEAGYVYIA 538

Query: 541 QPPIYGVKVGSEIKAYIQPGVGNQSEELRQALDYSSGRSKPTVQRYKGLGEMDDHQLMET 600  
 QPPIYGVKVGSEIK YIQPG+QE+L+AL YS GR+KPT+QRYKGLGEMDDHQLMET  
 Sbjct: 539 QPPIYGVKVGSEIKAYIQPGADQSEIKLQALARYSBGRKPTKPTQRYKGLGEMDDHQLMET 598

Query: 601 TMDPEHRLMARVSVDDAAEADKIFDMLAGDRVSPRRRFTSENAVYSNLDI 650  
 TMDPE+RLMARVSVDDAAEADKIFDMLAGDRVSPRRRFTSE NAVIS LD+  
 Sbjct: 599 TMDPEHRLMARVSVDDAAEADKIFDMLAGDRVSPRRRFTSENAVYSILDV 648

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1317> which encodes the amino acid  
 sequence <SEQ ID 1318>. Analysis of this protein sequence reveals the following:

Possible site: 41  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1698 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 584/650 (89%), Positives = 618/650 (94%)

Query: 1 MTEETQMSQRAQGYDASQIQVLEGLEAVMRPGMYIGSTKEGLAHLVWEIVDNGIDEA 60  
 M EE K+ E+ QSYDASQIQVLEGLEAVMRPGMYIGST+KEGLAHLVWEIVDNGIDEA  
 Sbjct: 1 MTEENKHFEKKMGYDASQIQVLEGLEAVMRPGMYIGSTAKEGLAHLVWEIVDNGIDEA 60

Query: 61 LAGFASHIKVYIEPINSITVVDDGRGIPVDIQKTRGPAVETVPTVLHAGKGFGGGGKVK 120  
 LAGFA HIKV+IE DNSITVVDDGRGIPVDIQ KTRGPAVETVPTVLHAGKGFGGGGKVK  
 Sbjct: 61 LAGFASHIKVPIEACNSITVVDDGRGIPVDIQKTRGPAVETVPTVLHAGKGFGGGGKVK 120

Query: 121 SGGLHGVGSSVVMALSTQGLDVRVYIKQIHYQEFKRGAVADLEVIQTTDVQTTVHTPT 180  
 SGGLHGVGSSVVMALSTQGLDVRVYIKQ+HYQE+RG VV DLE+IG TD+GTTVHTPT  
 Sbjct: 121 SGGLHGVGSSVVMALSTQGLDVRVYIKQIHYQEFKRGAVADLEVIQTTDVQTTVHTPT 180

Query: 181 DPEIFTEITTFDFDKLAKRIQELAPLNRGLRISISDKREGQVEKVEHYHGGIGSYVEFI 240  
 DPEIFTEITTFD+ LAKRIQELAPLNRGL+ISI+DKR G E E+ YEGGIGSYVEFI+  
 Sbjct: 181 DPEIFTEITTFQFYSLAKRIQELAPLNRGLKISITDKRSGMBQSEHFLYBGGIGSYVEFI 240

Query: 241 NENKEVIFENPIYTDGELDGISVEVAMQYITQYQETVMSFANNIHTHEGGTHQGFRAL 300  
 N+ K+VIFE PIYTDGEL+GI+VEVAMQYIT QYQETVMSFANNIHTHEGGTHQGFRAL  
 Sbjct: 241 NDKKDVFIPPIYTDGELGLIIVAMQYITTSYQYQETVMSFANNIHTHEGGTHQGFRAAL 300

Query: 301 TRVNDYAKNNKILKENEDNLTGEDVRGGLTAVISVKHPNPFQEGQTKIKLGNSVVKIT 360  
 TRVNDYAKNNKILKENEDNLTGEDVRGGLTAVISVKHPNPFQEGQTKIKLGNSVVKIT  
 Sbjct: 301 TRVNDYAKNNKILKENEDNLTGEDVRGGLTAVISVKHPNPFQEGQTKIKLGNSVVKIT 360

Query: 361 NRLFSEAPNRLFLENPQVAKIKVEKGLIASKARIAAKRAREVTRKSGLEISNLFKGLAD 420  
 NRLFSEAP RFLLENPQVAKIKVEKGLIASKARIAAKRAREVTRKSGLEISNLFKGLAD  
 Sbjct: 361 NRLFSEAPRFLLENPQVAKIKVEKGLIASKARIAAKRAREVTRKSGLEISNLFKGLAD 420

Query: 421 CSSNNAEMNELFIVEGDSAGSAGSGRNRFPAILPIRGKILANVEKATMDKILANEIRS 480  
 CSSN+A NRLFIVEGDSAGSAGSGRNRFPAILPIRGKILANVEKATMDKILANEIRS  
 Sbjct: 421 CSSNDRNQNELFIVEGDSAGSAGSGRNRFPAILPIRGKILANVEKATMDKILANEIRS 480

Query: 481 LFTAMGTGFGADFDVSKVRYQKLVIMTDADVDGAHIRTLLLTLLIYRMPRVLEAGYVYIA 540  
 LFTAMGTGFGADFDVSK RYQKLVIMTDADVDGAHIRTLLLTLLIYRMPRVLEAGYVYIA  
 Sbjct: 481 LFTAMGTGFGADFDVSKRYQKLVIMTDADVDGAHIRTLLLTLLIYRMPRVLEAGYVYIA 540

Query: 541 QPPIYGVKVGSEIKAYIQPGVGNQSEELRQALDYSSGRSKPTVQRYKGLGEMDDHQLMET 600  
 QPPIYGVKVGSEIK YIQPG+QE+L+AL YS GRSKPTVQRYKGLGEMDDHQLMET  
 Sbjct: 541 QPPIYGVKVGSEIKAYIQPGIDQEDQLKTSKYSYGRSKPTVQRYKGLGEMDDHQLMET 600

-512-

Query: 601 TMDPENRLMARVSVDDAARADKIFDMLMGDRVEPRDPFIEANAVYSNLDI 650  
 TMDPENRLMARVSVDDAARADKIFDMLMGDRVEPRR+FIE NAVYS LDI  
 Sbjct: 601 TMDPENRLMARVSVDDAARADKIFDMLMGDRVEPRDPFIEANAVYSTLDI 650

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 407

A DNA sequence (GBSx0442) was identified in *S. agalactiae* <SEQ ID 1319> which encodes the amino acid sequence <SEQ ID 1320>. Analysis of this protein sequence reveals the following:

Possible site: 26  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3186 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:GAA91552 GB:Z67740 unidentified [Streptococcus pneumoniae]  
 Identities = 82/142 (57%), Positives = 105/142 (73%)  
 Query: 45 LKESTADAIAYFIPESADFLKEYKANERAKVLETPILQGAKEKLLAKIQRQSRNPLVSHR 104  
 LK ST AI F P +FL++IK NEA+ LE PILF+G +LL I QS R+FLVSHR  
 Sbjct: 2 LKYSTPAIETFPANLENFLEKTYKNEARELEHPILFEGVSDDLEILNQSGSRNPLVSHR 61  
 Query: 105 DNOVIVILEKTEIIDYPTFVVTADNGSPSRKPSPESMYLKKEYQIDNCLVIGDRDIDKQK 164  
 ++QV+ ILEKT I YPTFVVT+ +CF RKP+PESMYL+EKVQI + LVIGDR ID +A  
 Sbjct: 62 NDQVLEILEKTSIARYFTEVVTSSGFKRKNPESMYYLREKYQISSGLVIGORPIDIEA 121  
 Query: 165 GBSAGFDTLLVDGSKSLMIEIE 186  
 G++AG DT L +L +++  
 Sbjct: 122 GQAAGLDTLPLTSIVNLKQVLD 143

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1321> which encodes the amino acid sequence <SEQ ID 1322>. Analysis of this protein sequence reveals the following:

Possible site: 59  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2472 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 122/185 (65%), Positives = 145/185 (77%)

Query: 1 MNYHYDIWDLGGTLLNLYESSTRAFPVETLKEFGYQAHDSVYQKLESTADAIAYFIPES 60  
 MNY DIWDLGGTLLNLYE ST+APV+TL F DHD+VYQKLESTADAIAYFIPES  
 Sbjct: 4 MNYQDIWDLGGTLLNLYESLTAQAFVQTLAPFSLPGCHDAVYQKLESTADAIAYFIPES 63  
 Query: 61 ADPLKEYKANERAKVLETPILQGAKEKLLAKIQRQSRNPLVSHRQNOVIVILEKTEIIDY 120  
 +FL Y+ EA L PI GAKE+L KI GSRNPL+SHRD CV +LE+ ++ Y  
 Sbjct: 64 PEPLHVYRLREADKLAQPIWCLGAKGKILGKIATSGSRNPLISHRDQCVNQLRQAGLLIY 123  
 Query: 121 FTEVVTADNGSPSRKPSPESMYLKKEYQIDNCLVIGDRDIDKQKBSAGFDTLLVDGSKS 180  
 FTEVVIA NGF+RKP+PES+ YLKEKY I++ LVIGDR IDKQAG++AGP+TLVDGSKS  
 Sbjct: 124 FTEVVTASNGFARKPNPESLYLKKEYQIDNGLVIGDRIDKQAGQAGPNTLLVDGRKN 183

-513-

Query: 181 LMEII 185  
L+EI+  
Sbjct: 184 LLSIV 188

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 408

A DNA sequence (GBSx0443) was identified in *Sagalactiae* <SEQ ID 1323> which encodes the amino acid sequence <SEQ ID 1324>. This protein is predicted to be stage V sporulation protein E (rodA).

- 10 Analysis of this protein sequence reveals the following:

Possible site: 42  
>>> Seems to have a cleavable N-term signal seq.  
INTEGRAL Likelihood = -11.15 Transmembrane 206 - 222 ( 177 - 226)  
INTEGRAL Likelihood = -10.14 Transmembrane 58 - 74 ( 50 - 82)  
15 INTEGRAL Likelihood = -9.34 Transmembrane 182 - 198 ( 177 - 205)  
INTEGRAL Likelihood = -8.55 Transmembrane 158 - 174 ( 156 - 177)  
INTEGRAL Likelihood = -8.12 Transmembrane 300 - 316 ( 299 - 324)  
INTEGRAL Likelihood = -2.66 Transmembrane 86 - 102 ( 83 - 102)  
20 INTEGRAL Likelihood = -2.34 Transmembrane 338 - 354 ( 338 - 357)  
----- Final Results -----  
bacterial membrane --- Certainty=0.5458 (Affirmative) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
25 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9669> which encodes amino acid sequence <SEQ ID 9670> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB15838 GB:299123 alternate gene name: ipa-42d-similar to  
cell-division protein [Bacillus subtilis]  
30 Identities = 142/392 (36%), Positives = 237/392 (60%), Gaps = 23/392 (5%)  
Query: 10 QKSNYPKQIDYAVVVPVFLMIGLASIYVA-TMDNPSNTIYAMPQQVSWIIMGICIA 68  
Q+S +++G D + VFF+ I + SIY A Y + +I QQ+ + ++G +  
35 Sbjct: 7 QQSPFQQG--DLIFIFGVFFI--ISVSVIYAAGQFGQGVNTDNI---QQIVFYLGLVAI 59  
Query: 69 FVVMLESTEFMHWATPYLYALGLTLMVLPILIFYSQLPAA--GAKNWTIQSVTLPQS 126  
V++ F E L K + Y++ +G+ +++ I SP+ A GAK+W IG +T+ QPS  
40 Sbjct: 60 TVLVLPDLEQLKLSLYPIFIQILSLIILKI--SPESIAPIVKGAKSWFIRITTI-QPS 116  
Query: 127 EFMKISYIIMLSRITVSFHQNRKTFQDDWKLL-GLPGLVTLFVMIIMLLQKDLGTALVF 185  
EFMK+ I+ML+ + + K +T +DD LL + G+ +PV ++LM +D GTA +  
45 Sbjct: 117 EFMKVLGIMMLASVIGKANPKQVTRDRDIDILLGLAGVAVIPVGLILM--QDAGTAGIC 174  
Query: 186 LAILSGILLISGISWMIILPILSTIVLEIASFLMIFISPNGKSWFYNLGMTYQINRLSA 245  
+ I+ ++ +SGI+W +I I + +L+L L++ I N + ++G+ YQI R++  
50 Sbjct: 175 MFVILNVFMESGINKWLLIALTAGSGILLISLILLVMI--NPPDVAKSVGIQVQLKRVTS 232  
Query: 246 WIDPFPSFAD---SIATQCTQMWISGSGGVTKGFNILELSPVRESMDIPTVIARNFGF 302  
W+ + + ++Q Q +++IGSGG+ G G +L+ VP +D IF++I E+PGF  
55 Sbjct: 233 WVSANETQSDSDNSGVDQDQALMAGSGGILNGCISNLKVVPESTTDFISITGESGPF 292  
Query: 303 IGSAIVIGLYLI I IYRMLRT--ITSNNQYTFYFISTGFMNIVHFVFNENIGAAWGILELT 360  
IG AIV+ ++ +IYR++ + I N+P +F G+ +IV H F+NIG +GI+P+T  
60 Sbjct: 293 IGCAIVVMFFFLIYRLVLDIKHPPNFRSFFCVGYTALIVLHTFQINIGNIGIMPVT 352  
Query: 361 GIPLPFISQGGSSLLSNLIGILWLSMSYQNT 392  
GIPL F+S GGSS LS LIG G+V + S Q T  
65 Sbjct: 353 GIPLEFVSXGSSSTLSTLIGRGIYVNASVQLT 384

There is also homology to SEQ ID 1028.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### 5 Example 409

A DNA sequence (GBSx0444) was identified in *S.agalactiae* <SEQ ID 1325> which encodes the amino acid sequence <SEQ ID 1326>. Analysis of this protein sequence reveals the following:

```

Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3195 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1327> which encodes the amino acid sequence <SEQ ID 1328>. Analysis of this protein sequence reveals the following:

```

Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2735 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 38/55 (69%), Positives = 48/55 (87%)

Query: 8 DEFKEAIDKGYISGNTVAIVRKNGKIFDYVLLHREVEVREBEVVTVERVLIDVLRLKS 62
 DEFK+AIID GYI+G+TVAIVRK+G+IFDYVL HE+V+ EVVT E+V +VL +LS
Sbjct: 5 DEFKQAIIDNGYIAGDTVAIVRKDGQIFDYVLPHEKVKNGEVVTKEKVEEVLVELS 59

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 410

A DNA sequence (GBSx0445) was identified in *S.agalactiae* <SEQ ID 1329> which encodes the amino acid sequence <SEQ ID 1330>. Analysis of this protein sequence reveals the following:

```

Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4241 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1331> which encodes the amino acid sequence <SEQ ID 1332>. Analysis of this protein sequence reveals the following:

-515-

Possible site: 41  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

5                   bacterial cytoplasm --- Certainty=0.4551(Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

10       Identities = 57/66 (86%), Positives = 63/66 (95%)

Query: 1 MSQEKLRSKLDQAKGSAKSGPGKITGDKLEAKGPFIEKTIANGKELADDADAVEGAVIDA 60

MS+EKLRK+QA GS KEG GK+TGDKLEAKGPFIEKTIANGKELADDADAVEGAVIDA

15       Sbjct: 1 MSEPRLKSKIBQASGGGLKPGAGKITGDKLEAKGPFVEKTIANGKELADDADAVEGAVIDA 60

Query: 61 VKEKLLK 66

VKEKLLK

Sbjct: 61 VKEKLLK 66

20       Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 411

A DNA sequence (GBSx0447) was identified in *S.galactiae* <SEQ ID 1333> which encodes the amino acid sequence <SEQ ID 1334>. This protein is predicted to be TnpA (orfB). Analysis of this protein

25       sequence reveals the following:

Possible site: 16  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

30                   bacterial cytoplasm --- Certainty=0.3961(Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9667> which encodes amino acid sequence <SEQ ID 9668>  
 35       was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1335> which encodes the amino acid sequence <SEQ ID 1336>. Analysis of this protein sequence reveals the following:

Possible site: 55  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

40                   bacterial cytoplasm --- Certainty=0.3365(Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45       An alignment of the GAS and GBS proteins is shown below:

Identities = 152/160 (95%), Positives = 154/160 (96%)

Query: 1 MKRMALPKGATVKKTKALKTKTKTYPCNLLNQKFNPDKNQWSTDTFTYISIGYKCYVYL 60

MKRMALPKGATVK KTKLK+TKTKTYPCNLLNQKFNPDKNQWSTDTFTYISIGYKCYVYL

50       Sbjct: 194 MKRMALPKGATVKKTKALKTKTKTYPCNLLNQKFNPDKNQWSTDTFTYISIGYKCYVYL 253

Query: 61 CAIIDLYSRKYLAWKLSHRMDAKLACDTLELALNKRKISGTLPHSDQGSQFKARERKI 120

CAI+DLYSRK LAKLSHRMDAKLACDTLELALNKRKISGTLPHSDQGSQFKARE RKI

55       Sbjct: 254 CAIIDLYSRKCTAWKLSHRMDAKLACDTLELALNKRKISGTLPHSDQGSQFKARELRKI 313

-516-

Query: 121 IDNNIMHSFSKPKRYPYDNAVTEAFFKYLAKRQINQKNYQ 160  
 IDNN IMHSFSKP YPYDNAVTEAFFKYLAKRQINQK YQ  
 Sbjct: 314 IDNTIMHSFSKPKGYPTNAVTEAFFKYLAKRQINQKRYQ 353

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 412

A DNA sequence (GBSx0448) was identified in *S. agalactiae* <SEQ ID 1337> which encodes the amino acid sequence <SEQ ID 1338>. Analysis of this protein sequence reveals the following:

10 Possible site: 27  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1090 (Affirmative) < succ>  
 15 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 413

A DNA sequence (GBSx0449) was identified in *S. agalactiae* <SEQ ID 1339> which encodes the amino acid sequence <SEQ ID 1340>. This protein is predicted to be histidine kinase (resE). Analysis of this protein sequence reveals the following:

25 Possible site: 40  
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -11.57 Transmembrane 17 - 33 ( 6 - 38)  
 INTEGRAL Likelihood = -4.67 Transmembrane 147 - 163 ( 142 - 166)

30 ----- Final Results -----  
 bacterial membrane --- Certainty=0.5628 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 35 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GF:AA25109 GB:AF140356 VncS [Streptococcus pneumoniae]  
 Identities = 178/435 (40%), Positives = 281/435 (63%), Gaps = 1/435 (0%)

40 Query: 1 MCKLAKIPKMFQIPIFSLQILVLSLFPFIFPFTYLETAKVKIHTMADRISKNMNGKR 60  
 MK+ +F K+PI PSI +L+Z +H +F+FP TYL R+ I A I++ ++ GK+  
 Sbjct: 1 MKRTGLFAKIPITFTSFISVLVLCILHATYFLFSTYLSHRQETTGQKATAIQSLGEGD 60

45 Query: 61 LKYLQDTLRLYSKSSDILKVFIRKNNRNKELQINDNNVNVKSDSNLSLIEREIKLHDK 120  
 + +Q L+LYS+ +SDIK +K +++L++ D++ ++ + SL IERE+K DG  
 Sbjct: 61 RQISQVLVDLYSQTSDIKGTIVKGMTEDKLEVKOSLPLTDRTQTSLSLIEREIVKTDGG 120

50 Query: 121 KHLQFVSTADMQKDAKDLKRLFLYSLSISFLPSIVLSLIYAKSIKNNIQRITMVTDM 180  
 + LQF+++ D+OK+A+ +SL+FLPY+L SFL S++++ IYA++I I RI VT +M  
 Sbjct: 121 TWILQFLASMDLQKBAQISLQFLPYLLASFLISLIVAYIVARTIVAPILEIKRVTIRM 180

Query: 181 IKDLKELTKLSSNDKIQGLKQKQINDLYCALNTINDLEPKNEIKLGLKLYDFFPKAS 240  
 + LD + RL++ S DEIG LK+QIN LY LL I DL KN+ IL+LEK+K +F +GMS  
 Sbjct: 181 MDLDSQVLRVDSKDEIQNLKQKQINSYQLLLTVIADLHEKRAILQLEKMKVFLRGS 240

-517-

Query: 241 HELKTPFLSLKILLERMKYNIQYKDRDFEYISCTINIVMLTKVQSILSFYSIKDLAND 300  
 HELKTPFLSLKILLERMKYNIQYKDRD Y+ + IVD L +V QILS S++L +D  
 Sbjct: 241 HELKTPFLSLKILLERMKYNIQYKDRDQYLGVALGIVDELNHHVQLILSLSVQELRDD 300

5 Query: 301 EEVLAVPGDILDEVELEKYSILVWQKIKINKELLDNIIYIKRTAINIVPSNLISNAVKYITN 360  
 E +++ +++ Y+L L +++ I+ L Y+ + + + SMLISNAVK+K+  
 Sbjct: 301 RETIDLLQMTQNLNKKVYALLAKERELQIDNSLTHQQAYINPSVWKLILSLISNAIKHV 360

10 Query: 361 RRGITNIIKIANWLLIENSVDKNIKIKINKILDASFDKLINSNGICLNIVENILNKYNI 420  
 G++ I L IENS + K+ + + K+ S G+GL +VK+L +  
 Sbjct: 361 PGGLVRIERGEELFIENS CSSEBQEKLAQSFDNASKRVKGS -GMGLFVVKSLLEHEKL 419

Query: 421 KYEILHGENYPTFKI 435  
 Y EN F I  
 Sbjct: 420 AYRFEMEENSITFFI 434

15

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1341> which encodes the amino acid sequence <SEQ ID 1342>. Analysis of this protein sequence reveals the following:

20 Possible site: 37

>>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -11.83 Transmembrane 14 - 30 ( 6 - 35)  
 INTEGRAL Likelihood = -2.44 Transmembrane 157 - 173 ( 156 - 174)

25 ----- Final Results -----  
 bacterial membrane --- Certainty=0.5734 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

30

The protein has homology with the following sequences in the databases:

>GP:AAD25109 GB:AF140356 VncS [Streptococcus pneumoniae]  
 Identities = 123/455 (27%), Positives = 223/455 (48%), Gaps = 23/455 (5%)

35 Query: 3 LIKKTFLVINGLIVVVTISILLVLYFAMPIYTKVKDKVKCEPDQTSKIKGTVEIR 62  
 L K F+ + V+V + L +YF P Y + + + + + ++ +GK I  
 Sbjct: 6 LFAKFIYITTSIFSVLVICHLAIYFLPSPSTYLSHRQETIQKATAIQSLGSKDRQSIG 65

40 Query: 63 DILTCKKINKNIKYSLVDSNDQLLYPSLQLLDGVSESKDSQNNIVITTFNSYSNVKNWS 122  
 +L +I ++ ++ L++ D + D Q ++ +  
 Sbjct: 66 QVLDLYSQTSIDIKGV--KGMETDKLKVKDSLPLOTDRQTTSLF-----IEE 111

Query: 123 QKVTLRDGGKMTLLQCSLQFVTDASKVLLDLYPSLLIFSVTVGSIVAYLSRTSSRIL 182  
 ++V +DG M L +S+ +A ++ L P L+ S + +VAY+Y+RT IL  
 45 Sbjct: 112 REVKTQDGGTMIQLFLASMDLQEARQLSLOFLFYTLASFLSLVAYIYARTIVAPIL 171

Query: 183 SMSQTAKKMWLEPNLTCTHKGDEIAMLASDINRLYASLSSTIKSLQKYEKASDSERR 242  
 + + ++M++L+ + KDEI L IN LY L T I L ++ E E+  
 Sbjct: 172 EIKRVITRRMDLQVRLRVDSKDEIGNLCKQINSLYQHLVITADLHENEAILQLEK 231

50 Query: 243 KSEFLRMTHSLKTPITSVIGMDGLNVDGPDADRKYLRKCRDVLBOQAQVLSILSL 302  
 K EFLR SHELKTP+ S+ +I+ M +G + DRD+YL +++ V ILSL  
 Sbjct: 232 KVEFLNGASHELKTPFLASLKILLENRMENIRYKDRDQYLGVALGIVDELNHHVQLILSL 291

55 Query: 303 SKIETLASQNELFLSKSLEEDMEVFLVLSLKHLLKVTINLSQEPVQKANKVYLLKATKN 362  
 S ++ L ++E L + ++ + +L++ + L++ +L Q N + + N  
 Sbjct: 292 SVQVEL-RDRETFIDLLQMTQNLNKKVYALLAKERELQIDNSLTHQQAYINPSVWKLILSN 350

Query: 363 IIDNAPHYTKSGSQVMQLKDNOLVIKNEARTLLTQQQMKLPQPPYRPPYSRNRKDDGT 422  
 -I NA ++ G V I ++ +L I+N + ++ ++L Q F + +RK G+  
 60 Sbjct: 351 LISNAIKHVSFGLVRIERGEELFIENSC---SSBQEKLAQSF---SDNASKRVKGS 403

Query: 423 GLGLFITHQLDQHHLAYRFVVLDDQRNMVPTIDFP 457  
 G+GLF+ +L+ LAYRF +++ + F IDFP  
 65 Sbjct: 404 GMGLFVVKSLLEHEKLAYRF-EMEENSITFFIDFP 437

An alignment of the GAS and GBS proteins is shown below:

Identities = 108/454 (23%), Positives = 220/454 (47%), Gaps = 22/454 (4%)

```

5 Query: 4 LKIPFFMFQIPISIGLIIIVHSIIFFPKTYLTKRVKIHIMADEISKMMNGKELKY 63
 +++ K F + I ++ +++ + + +F P Y + + + D+ SK + GK +
 Sbjct: 1 VRLIKTKTFLVINGLIIIVVTSIILVLVFPMPYTYTVKDKRVKCEFDQTSKQINGKIVTE 60

10 Query: 64 LDQTLELYSKSSDKVFKKNNK-----NEIQINMINMNVVKSNSN--SIIT 109
 + L + I + ++N+ +E + + N+N+ D++ ++ +
 Sbjct: 61 IRDILTKKINKINWISLVDSNQQLYPSQLLDGVSESKDSQNVNIVTVTDSNYSNVK 120

 Query: 110 ERREIKLHDGKKIHLQFVSTADMQKDAKLSLKLFLPSLSISLFSIVISLIYAKSIKN 169
 +++ L DKK+ L S+ DA + L P L S +++ +Y+++
 Sbjct: 121 MSQKVLTRDGGKMTLLGQSSLQPVTDASKVLLDLYPSLLIFSVTGSIYAYLYSRSSRR 180

 Query: 170 IQEITMVTDMKIKLDKETRLKISSNDEIGQLQKINDLYCALLMTINDLEFNKEILKE 229
 I ++ KM+ L+ I DEI L IN LY +L +I L+ ++ + E
 Sbjct: 181 ILSMSQAKKMVNLENLCTITNGKDEIAMLASDINRLYASLSTSISLQKEYEKASDSE 240

20 Query: 230 KIKYDFPKGASHELTPLSSLKILLEMKYNIGKYKDRDPYISBCINIVNLTYSQVSL 289
 + K +F + SHELKTP++S+ +++ M YN+G + DRD Y+ +C ++++ + V IL
 Sbjct: 241 REKSEFLRMTSHELTPTITSVIGIMIDMLYNVGDPAADRKYLRKCRDVLGGQQLVQSIL 300

 Query: 290 SFYSIKDL-NNDEEYLVNGVDLDEVLEKISILVNQKKINKINKELLDNIYIGKTALAVF 348
 S I+ L + ++E ++ +L+E +E +L K + + L + + K L
 Sbjct: 301 SLSKIETLASQNLQELFSLKSLSEBEKEVFLVLSKELKLVTLNEBQPVKANKVYLKAI 360

 Query: 349 SNLISNAVKYTRNGIINKIANDMLIENSVDKWKISKINKILDASF-----DLKLN 402
 N+I NA TT G + I+ ++ I+IN + + K L F + D
 Sbjct: 361 KNIIIDNAFHYTSGGQVHLQKDMVLVIRNEAETLLTQQQKQLPQFFYRPDYSKRNKG 420

 Query: 403 SNGLGLNIVKHILNKVNIKYB-ILGHENYPIFKI 435
 GLGL I II++++Y Y ++ + +F I
 Sbjct: 421 GTGLGLPITHQILDGLHAYRPFVVDQRWVFYTI 454

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 414

- 40 A DNA sequence (GBSx0450) was identified in *S.agalactiae* <SEQ ID 1343> which encodes the amino acid sequence <SEQ ID 1344>. This protein is predicted to be response regulator (regX3). Analysis of this protein sequence reveals the following:

```

Possible site: 34
>>> Seems to have no N-terminal signal sequence
45 INTEGRAL Likelihood = -0.80 Transmembrane 50 - 66 (50 - 66)

----- Final Results -----
 bacterial membrane --- Certainty=0.1319 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
50 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9665> which encodes amino acid sequence <SEQ ID 9666> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

- 55 >GP:AAD25108 GB:AF140356 VncR [Streptococcus pneumoniae]  
 Identities = 131/218 (60%), Positives = 176/218 (80%), Gaps = 1/218 (0%)
- Query: 5 MKILTVEDDKLIRGISEYLSFPGYTVIQAKDCRRALSKFNS-DINLVILDIQIPFINGL 63



-519-

1                   MKIL VED+++IREG+S+YL++ GY I+A DG+EAL +F+S ++ LV+LDIQ+P +NKL  
 Subjct: 1       MKILIVEDDEMIREGVSDFLTDCGYETLEADQQAALBQFSSYEVALVLDIQMPKLEKL 60  
 5                   Query: 64   EVLKEIRKKGNLFLILITAFSDSEYKIDAFINLVGYVEKFPFLVFLVKARIDSLIKRMP 123  
                   EVI: EIRK S +P+L+ITAF DEEYK+ AF ++ DGY+EKFPFL +LK R+D++ K+ +  
 Subjct: 61       EVLAEIRKTSQVFFVLMITAFQDEEYMSAFASLADGYLEKFPFLSLKRVVDIAFKRYD 120  
 10                   Query: 124   HLEKTEYHNSLVNENSYTAKINDEKIDVNAKLEILKCLLMDQQLVTRMQIIDYVWKS 183  
                   F YE+ V+P SY+A + +++ +NAKLEIL L+ H+G+ DFR QID VWK +  
 Subjct: 121      TGRIFSYKDTKVDYFESYGASLQGVFPAKLEILDLVLKMGKALTRSQIIDAVKAT 180  
 15                   Query: 184   EIPYDRVVDVYIKELRKLLDCLITTRNVGKLERK 221  
                   +S+P+DRV+DVYIKELRKLL LDCI T+RNVGKLERK  
 Subjct: 181      DEVFPDRVVDVYIKELRKLLDCLITTRNVGKLERK 218

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1345> which encodes the amino acid sequence <SEQ ID 1346>. Analysis of this protein sequence reveals the following:

Possible site: 60  
 >>> Seems to have no N-terminal signal sequence  
 20           INTEGRAL   Likelihood = -2.60   Transmembrane   48 - 64 ( 48 - 64)  
 ----- Final Results -----  
                   bacterial membrane --- Certainty=0.2041(Affirmative) < succ>  
 25                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAF72358 GB:AF192329 VanRB [Enterococcus faecalis]  
 Identities = 88/215 (40%), Positives = 128/215 (58%), Gaps = 2/215 (0%)  
 30                   Query: 3   KILVVEDDDTISQVCEPLKANNYDFDCVFGQAALDKWQTSYDLIILDIMPLSGLGE 62  
                   +IL+VEDDD I + FL Y D DG A K+ +Y L+ILDIMLP +G E  
 Subjct: 4       RILLVEDDDHICNTVRGFLAERAGYQVDACTQNEATYKFYEMTYQLVILDMPLFGNGSE 63  
 35                   Query: 63   VLKTIKRTSDVPIIMLTALDDEYTLQVSPNHLISDYVTKPFPSPILIKRIENVLVRSTFD 122  
                   +L+ R +D P+M+TAL D+ Q+ +F+ DYVTKPF IL+R+S +R S  
 Subjct: 64      LLRFRANKDTPILMTALSDDENQIRAFDAEDDYVTKPFPQIILLRVALLRRSGAL 123  
 40                   Query: 123   EKR-QIGDLLVDTEHSHVWQSTLVKLTKEVDIIDLAKRHQKIVTRDQLMDIDWYGS- 180  
                   K +G L + + +V GT +LT+KE++I+ L + + +T + + IWGY  
 Subjct: 124      AKEIRVGRIILLPDDFTVLCDDTELPTRKEPFIILLVQNGKRTILTHEILSRWGYDF 183  
 45                   Query: 181   ELDTRVLNHNKILNRKMGYIPLKTIITMGVYLGE 215  
                   E D + HIKNL R K+ +KTI G+GY L E  
 Subjct: 184      EGDGSVTHIKILNRKALPENIYKTIIRGVGYRLGE 218

An alignment of the GAS and GBS proteins is shown below:

Identities = 80/214 (37%), Positives = 126/214 (58%), Gaps = 4/214 (1%)  
 50                   Query: 6   KILTVEDDKLIREGISYLSREGYTVIQAKDQREALSQFNS-DINIVITDIQIPFINGLE 64  
                   KIL VEDD I + I E+L Y DG+ AL K+ + +L+ILDI +P +GLE  
 Subjct: 3       KILVVEDDDTISQVCEPLKANNYDFDCVFGQAALDKWQTSYDLIILDIMPLSGLGE 62  
 55                   Query: 65   VLKTEIRKSNLFLILITAFSDSEYKIDAFINLVGYVEKFPFLVFLVKARIDSLIKKFNH 124  
                   VLK IRK S++PI+LTA DE ++ +P +L+ YV KPF +L RI++++ +  
 Subjct: 63      VLKTIKRTSDVPIIMLTALDDEYTLQVSPNHLISDYVTKPFPSPILIKRIENVLVRSTFD 122  
 60                   Query: 125   LEKPEYHNSLVNENSYTAKINDEKIDVNAKLEILKCLLMDQQLVTRMQIIDYVWKS 184  
                   EK + +L V+ ++ ++ + KE +L+ L + + + +R Q+D +W SE  
 Subjct: 123      -EKRQIGDLLVDTEHSHVWQSTLVKLTKEVDIIDLAKRHQKIVTRDQLMDIDWYGS 181  
 65                   Query: 185   EIPYDRVVDVYIKELRKLLDCLITTRNVGKYL 218  
                   RV+D +TK LRKK+ + TI +GY L  
 Subjct: 182      --LDTRVLNHNKILNRKMGYIPLKTIITMGVYL 213

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 415

- 5 A DNA sequence (GBSx0451) was identified in *S.galactiae* <SEQ ID 1347> which encodes the amino acid sequence <SEQ ID 1348>. This protein is predicted to be Vexp3. Analysis of this protein sequence reveals the following:

Possible site: 49

>>> Seems to have an uncleavable N-term signal seq

10 INTEGRAL Likelihood = -12.68 Transmembrane 423 - 439 ( 413 - 447)  
 INTEGRAL Likelihood = -10.67 Transmembrane 16 - 32 ( 12 - 37)  
 INTEGRAL Likelihood = -9.77 Transmembrane 303 - 319 ( 301 - 326)  
 INTEGRAL Likelihood = -3.13 Transmembrane 343 - 359 ( 343 - 367)

15 ----- Final Results -----

bacterial membrane --- Certainty=0.6074 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 20 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD47594 GB:AF140784 Vexp3 [Streptococcus pneumoniae]

Identities = 280/458 (61%), Positives = 363/458 (79%), Gaps = 3/458 (0%)

25 Query: 1 MHNAPAVYTRKSLKLIILWILSNATLSIISLSIKDATDRASKETFMNTNSPFSMEIN 60  
 M+ NAFAPVYTRK KS+I L+IL NA+LS++ LSIK AT +AG+ETP NITNSPFSM+IN  
 Sbjct: 1 MLHNAFAYVTRKFPKSIIVTFLIILLNASLSVGLSIKATAKASQETFKNTNSPFSQIN 60

Query: 61 RQVNPGRTPGGGNGVKGDIKKISQNSIDSIVYKRINSVADLVHDHIIETQDTLANQSPER 120  
 R+VN GTPRG GN+KGSDIKCI++ +I+SYVGRIN++ DL +D+IET +T N + +R  
 30 Sbjct: 61 RRVNQOTPRGAGNIGRDIKKITENKALISYVVKRINAGULTGYDILLETFTKKNLTAOR 120

Query: 121 AKNFKRTVMITGVNDSAKETKPVSEAYKLVEGHLENKDKNKILMHKDLAKNNKLVGDK 180  
 AK F ++M+TGVNDS+KE KPVSEAYKLVEGHLENKDKNKILMHKDLAKNNKLVGDK  
 35 Sbjct: 121 AKRFGSSLAITGVNDSKEDKFPVSGSYKLVEGHEIINDDKKILLHKDLAAKHGKLVGDK 180

Query: 181 IKIKSNLPDADNEKVANETVVEIKGLFDGHNSSGGVSAQELVENTLITDVHSAKVVYN 240  
 +K SN++DADNEK A ETVEV IKGLFDGHN V+ +QELVENT ITD+H+AAK+YG  
 40 Sbjct: 181 VKLDSNLYDADNEKGAETVVEITIKGLFDGHNKSAVTYSQELVENTAITDHTAAKLYGY 240

Query: 241 TEDTAVYQDATFFVKGDKNLSDVIKDL-GKLDINWREYNLIKSSSNYPALQSSISGIYSI 299  
 TEDTA+Y DATTFV DNLD V+K+L G INW+ Y L+KSSSNYPAL+QSSIG+Y +  
 45 Sbjct: 241 TEDTAIYGDATFFVTADGNLSDVMKELNGISGINWKSAYTLVWSSSNYPALQSSISGMYN 300

Query: 300 SNLKPQSLPFGPVVSVLLFLMNNARKKEIAVLSSLSIKLETFPOOFIIMVFIIPAL 359  
 +N LP GSD F+ +++L LL LN+NR+KE+ +LSE+G+ + I OQFI E + I+IPAL  
 50 Sbjct: 301 ANLLFPQSLSPSVLLALLLGLMNNARKKEVGGILLSTGLAQSLIQQFTPSILAIIPAL 360

Query: 360 LGSYFLAQYTAOKLGNILNKVTGDIARQASASSQGGGAABGPNKTLSCGLDINV- 418  
 + +YFIA YTA +GN +L VT +AKQ ++ + +S LGGGAE +GF+KTLIS LDI++  
 55 Sbjct: 361 VSAYFLANYTARIGNTVLNVTSVAKQASAKAQAASNLGGGAENVDFGFTLSSLDISIQ 420

Query: 419 LPKFIYVVFVFPSPVLLVSLILSSIYTLRKNPKELLID 456

FII V+ + V+LV + L+S LRK PKELL+D

Sbjct: 421 TSDFIIFVIALVVLV-LV-MALASSNLRKPKELLID 457

55

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1349> which encodes the amino acid sequence <SEQ ID 1350>. Analysis of this protein sequence reveals the following:

Possible site: 31

>>> Seems to have no N-terminal signal sequence

-521-

|          |                     |               |                        |
|----------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -12.90 | Transmembrane | 19 - 35 ( 16 - 43)     |
| INTEGRAL | Likelihood = -7.27  | Transmembrane | 371 - 387 ( 359 - 392) |
| INTEGRAL | Likelihood = -7.01  | Transmembrane | 335 - 351 ( 326 - 357) |
| INTEGRAL | Likelihood = -6.21  | Transmembrane | 282 - 298 ( 276 - 308) |

5

----- Final Results -----

bacterial membrane --- Certainty=0.6158 (Affirmative) < succs  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succs  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succs

10

The protein has homology with the following sequences in the databases:

>GP: AAC24912 GB: AF012285 YknZ [Bacillus subtilis]  
 Identities = 176/408 (43%), Positives = 250/408 (61%), Gaps = 16/408 (3%)

15

Query: 1 MNNKFPALSSIWGHMRSIIITMLGIIIGVAIVIMGLNANMNSVTSSKQKDIQLY 60  
 +EN + ALSS+ HKMRSIIITMLGIIIGV +V+++ +G + + + S ++LY  
 Sbjct: 4 LENIRVALSSVLARHMRSIIITMLGIIIGVSVVVAVVQGGQKQLQSIQSPQNTVELY 63

20

Query: 61 PQEKGRS--EDLYAGLHTHNNHVKFPMLEQIVKIDIDGDSYFINSATSTISYEKKV 118  
 + EE + A + + ++K +K +GI + S + Y ++  
 Sbjct: 64 YMSDEELASNPFAAABSFTFENDIK-----LKGIBQIKQVASTSESMKARYHEET 117

25

Query: 119 DNASIIIGVSKDYFNKIVYDVGRTLTINDYNSNFSRIILLDTVLADLPFGKNYKALNK 178  
 D A++ G++ Y N+ + I +GCT TDND+ +R+ ++ +A +LF K S L +  
 Sbjct: 118 D-ATVNGINDGYMNVSLKIESGRTPTINDFLAGNRVGIIISQMAKELFDK---TSPLGE 173

30

Query: 179 VVSLDKDYLVIGVYKIDQTPVSPDLGGCAVMANTQVASEPQTEIGSIYIHVDIQS 238  
 VV ++ + +IGV K + +SFD LS V N + S FGT + ++ + V +  
 Sbjct: 174 VVMIGQPVEIIGLVKKVITGLLSFD-LSMIVFFN-HMKSSPQTSDFSNVSLQVESADDI 231

35

Query: 239 MNLGQAADMLTNISHIKDQYAVPNSKIVEEINQSFIMTTVIGSIAAILLVGGIGV 298  
 + G +AA L N +H + Y V + + I I +IMTT+IGSIA ISLLVGGIGV  
 Sbjct: 232 KSGAKAAQ-LVNDNIGTSDSYQNMWEEIAAGIGKVIAMTTIIGSIAISLLVGGIGV 290

40

Query: 299 MNIMLVSVTERTREIGRLKALGATRLKILSQPLIESVVLTVLQGLIGLLAQLSVGLGN 358  
 MNIMLVSVTERTREIG+RK+IGATR +IL+QPLIESVVLTV++GGI+G+ + A+ +  
 Sbjct: 291 MNIMLVSVTERTREIGIRKSLGATRQQLITQPLIESVVLTVLQGLVIGIG- YGGAALVS 349

45

Query: 359 AMTLKACISLDVALIANVLFASIGVFFGMLPANKASKLPDIALRYE 406  
 A+ + IS V VLFPS IGV PGMLEPANKA+KLPDIALRYE  
 Sbjct: 350 ALAGMPLISNQVCGSVLFSLMIGVIGMLPANKA+KLPDIALRYE 397

An alignment of the GAS and GBS proteins is shown below:

Identities = 56/247 (22%), Positives = 101/247 (40%), Gaps = 42/247 (17%)

45

Query: 147 YKLVEGKHLENKKNKI-----LMKDLAKKNLK-----VGDKIKISNLFDA 190  
 Y +V G+ L + D + ++ DL K N K + DK + ++  
 Sbjct: 136 YDIVAGRTLTINDYNSNFSRIILLDTVLADLPFGKNYKALNKVSLSDKDLVIGVYKKT 195

50

Query: 191 DNEKVANETVEVEIKGLFDGNSGGVSAAQELVENTLITD/HSAKVVYGNEDTAVYQDA 250  
 D V+ FDG + G V A NT + A +G E ++Y  
 Sbjct: 196 DQTPVS-----FDLGGCAVMA-----NTQV-----ASEPQTEIGSIYIHV 232

55

Query: 251 TFFVVGDKNLDVSKDL--GKLDINRVEYMLKSSSNYPALQSQISQIYVSNKLVSLG 308  
 ++ NL + D+ I +Y + +S + S + ++ + +G+  
 Sbjct: 233 ND-IQNSMNLGQAADMLTNISHIKDQYAVPNSKIVEEINQSFIMTTVIGSIAAIS 291

60

Query: 309 IFAGVVVSLFLPAGNARIGITAVLSLGSISLEHIFQPIIMVPSIIPALLGSYFLAQ 368  
 + G+ V ++ + + R +KI + +LG +A+I QF+IE V ++ + L LAQ  
 Sbjct: 292 LVGGIGVMNIMLVSVTERTREIGRLKALGATRLKILSQPLIESVVLTVLQGLIGLLAQ 351

65

Query: 369 TADKLGN 375  
 + LGN  
 Sbjct: 352 SVGLGN 358

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 416

A DNA sequence (GBSx0452) was identified in *S.agalactiae* <SEQ ID 1351> which encodes the amino acid sequence <SEQ ID 1352>. This protein is predicted to be Vexp2 (b0879). Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3194 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD47593 GB:AF140784 Vexp2 [Streptococcus pneumoniae]
Identities = 142/207 (68%), Positives = 169/207 (81%)

Query: 1 MDILEIKNVNYSYANSKEKVLGSGVQKFKFYAIVGKSGTGKSTLLSLLAGLDKVTG 60
 M +L++++V Y Y N+ E VL +N FE GKPY+I+G+SG GKSTLLSLLAGLD G
Sbjct: 1 MTLLQLQVDTYRYNTAFAVLYQINYNFEPKPYFIIIGSGAGKETLSLLAGLDSEVGG 60

Query: 61 KILFKNEDIEKQYSNHRKNNISLVFQNYNLIDYLSPIENIRLVNKSVDSEILFELGLDK 120
 IL+ EDI KKGYS HR ++ISLVFQNYNLIDYLSPIENIRLVN ++ L ELGLD+
Sbjct: 61 SILFQGEDIRKQYSYHRMHISLVFQNYNLIDYLSPIENIRLVNKGASNTLLGLDSE 120

Query: 121 KQIKRNVNKLGGGQQQRVAIARALVSDAPIILADEPTGNLDSVTAGEINILKELAQDRN 180
 QIKRNV++LSGGQQQRVAIAR+LVS+AP+ILADEPTGNLD TAG+I+ +LK LAQ
Sbjct: 121 SQIKRNVQLSGGQQQRVAIARSLVSEAFVILADEPTGNLDPKTAGDVELLKSIAQRTG 180

Query: 181 KCVIVVTHSKEVADSADITILESGKKL 207
 KCVIVVTHSKEVA ++DI LEL KKL
Sbjct: 181 KCVIVVTHSKEVAQASDITILEKDKKL 207
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1353> which encodes the amino acid sequence <SEQ ID 1354>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2717 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 83/230 (36%), Positives = 135/230 (58%), Gaps = 13/230 (5%)

Query: 1 MDILEIKNVNYSYANSKEKVLGSGVQKFKEL--GKPYAIVGKSGTGKSTLLSLLAGLDKVTG 58
 M +E+K V+ SY + V + FE+ G+ I+G SG GKST+L++L G+D V
Sbjct: 5 MDILEIKVQKSKSYQIGETVFPANHEVSFENNGELVVILGASGAGKSTVNLNLGGMTVD 64

Query: 59 TGKILFKNEDIE---KKGYSNHRKNNISLVFQNYNLIDYLSPIENIRLVNKSVDSE---- 111
 G+++ +DI K + +R+N I VFQ YNL+ D+ E+ L + V ++
Sbjct: 65 AGCVIIGDKGLAHYTSKALTQYRRNAIGFVQFPYLVNLTAENVLAVEIVADALDFV 124

Query: 112 -ILFELGLDKKQIKRNVNKLGGGQQQRVAIARALVSDAPIILADEPTGNLDSVTAGEIN 170
 IL ELGL + + +LSGG+QQRV+IARAL + +L DEPTG LD T +I+
Sbjct: 125 TILKEVGLSHR-LDHPFALSGGQQQRVSIARALAKNPKLLLCDEPTGALDYQTGRQILT 183
```

-523-

Query: 171 ILKELAQDRNKCIVVTHSKEVAD&ADIILELSGKKLKK--VNHQMLEVE 218  
 +L++MQ + V++VTH+ +A AD ++ + ++ K +NK +E  
 Sbjct: 184 LLQMAQTKGTFVVIVTHNPAIAPLDERVIFMHDQVTKTVINKEPASIE 233

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 417

A DNA sequence (GBSx0453) was identified in *S.agalactiae* <SEQ ID 1355> which encodes the amino acid sequence <SEQ ID 1356>. Analysis of this protein sequence reveals the following:

Possible site: 25  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -3.35 Transmembrane 17 - 33 ( 17 - 34)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.2338(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 418

A DNA sequence (GBSx0454) was identified in *S.agalactiae* <SEQ ID 1357> which encodes the amino acid sequence <SEQ ID 1358>. This protein is predicted to be Vexpl. Analysis of this protein sequence reveals the following:

Possible site: 56  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -11.52 Transmembrane 294 - 310 ( 285 - 312)  
 INTEGRAL Likelihood = -10.67 Transmembrane 396 - 412 ( 385 - 417)  
 INTEGRAL Likelihood = -8.76 Transmembrane 17 - 33 ( 14 - 38)  
 INTEGRAL Likelihood = -4.14 Transmembrane 335 - 351 ( 333 - 357)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.5607(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AD47592 GB:AF140784 Vexpl [Streptococcus pneumoniae]  
 Identities = 165/425 (38%), Positives = 271/425 (62%), Gaps = 4/425 (0%)  
 Query: 2 IKVALIAYITRKKNRTLIIFAILITVLSCLYSCITIMKSSNRIKALYESNNSISITK-K 60  
 I+ + AY+RK+ R+ I+F IL ++L+ + +CLT+MKG+ +E LY+S N+S SI K +  
 Sbjct: 4 IQRSWAYVSRRKRLRSPILPLILLVLLAGISACLTIMKSNKTVESNLKYSINTSPSIIKIE 63  
 Query: 61 DGKYPNINQFFNIKIKIKVEEKTIPQYDGLAKLDKLVSGRQSNREDELSDFKNNVSL 120  
 +G+ F ++ ++ KIK +E + + +AKLKD + V+GEQS+ R+DLS N+VSL  
 Sbjct: 64 NGCTPKLSDLASVSKIKGLNVSPFLKTVAKLKDKKRAVTGEQSVERRDLSAANDNLVSLT 123  
 Query: 121 ATSNPTKRNLLPSSGVSPFKEGGNIEKNDKNIILVHKFPAKQNKLKGLGEIDLELLUTSKS 180  
 A ++ +++ F+S F+ KEG++++ D IL+HEE AK+N L L D+I L+ +E S  
 Sbjct: 124 ALDESGRDVTFPTSSAFNLKGRHLQKGDSEKILIHRELAKNGLSLHKQIGLDAQSE-S 182

-524-

Query: 181 GKIKSHKFKIIGIFSGKKQETTTGLSSDFSENMVFDYSTSQRLINKSENNRIANKIMLY 240  
 GK ++ +F+IIGIFSGKKQK +TGLSSDFSENV VF DY +SQ +L SR A + Y  
 5 Sbjet: 183 GKQIVSEFIIIGIFSGKKQKFTGLSSDFSENVFDYTSQRLINKSEQAQVBAARF -Y 240

Query: 241 SGLSESTELALNKLDKFKIDKSKYSIKKDKAFESLESVSGIKHIIKIMTYSIMLGIV 300  
 + + + + + + + + + + Y +K+NKAF + +SV+ + + I Y + + G  
 10 Sbjet: 241 VENPKENDGLMKQVENLALENQGVKSEKAFBQIKDSVATFTQTLTIFLYGMLLAGAG 300

Query: 301 VLSLILILMLRERIYIGIFLISQTTKIQIIRQFIFELIFISIPSISSLFLGNLLLAIV 360  
 L L+L LMLRER+YE+GI L+G K I QF B++ +S+ + + + + GN + +  
 10 Sbjet: 301 ALILVLSLMLRERIVYGILLALGKRSKIFLOFCLFVVLNVLGALLPAFVAGNAITTYL 360

Query: 361 VEGFINSSENSMIFGSLINKSSFMNITTLAESYILILISIVLVMVASSLLEPKFKEI 420  
 ++ + S + +L SS +I +AESY+ L+ + LSV + + K PKEI  
 15 Sbjet: 361 LQTLASGDQASLQDTLAKASSLSTLSLFAESYVFLVLLSCLSVLCLFLFKRSKPEI 420

Query: 421 LSKIS 425  
 LS IS  
 20 Sbjet: 421 LSSIS 425

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1359> which encodes the amino acid sequence <SEQ ID 1360>. Analysis of this protein sequence reveals the following:

Possible site: 15  
 25 >>> Seems to have an uncleavable N-term signal seq

|          |            |          |               |                        |
|----------|------------|----------|---------------|------------------------|
| INTEGRAL | Likelihood | = -11.57 | Transmembrane | 23 - 39 ( 16 - 43)     |
| INTEGRAL | Likelihood | = -11.36 | Transmembrane | 371 - 387 ( 362 - 396) |
| INTEGRAL | Likelihood | = -9.12  | Transmembrane | 331 - 347 ( 324 - 360) |
| INTEGRAL | Likelihood | = -7.70  | Transmembrane | 280 - 296 ( 277 - 308) |

30 ----- Final Results -----

|                     |     |                               |         |
|---------------------|-----|-------------------------------|---------|
| bacterial membrane  | --- | Certainty=0.5628(Affirmative) | < succ> |
| bacterial outside   | --- | Certainty=0.0000(Not Clear)   | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000(Not Clear)   | < succ> |

35

The protein has homology with the following sequences in the databases:

>GP:AA97962 GB:U96166 ATP-binding cassette transporter-like protein  
 [Streptococcus cristatus]  
 Identities = 222/311 (71%), Positives = 278/311 (89%)

40 Query: 16 MRSILTMLGIIIGIGAITAIFSIIEGNTENTKRLQIGSSNNTIVFNKSSSIDPKPPK 75  
 MRS+LWMLGIIIGIGAITAIFSIIEGNTENTKRLQIGSSNNTI +V+KKS+IDP P+K  
 Sbjet: 1 MRSMLTMLGIIIGIGAITAIFSIIEGNTENTKRLQIGSSNNTIKVYDKKS+IDPSIPEK 60

45 Query: 76 SNAKKPDYLPMAHELSKIQQVQVGNALISYGDIDKVTYHLQKSSAKISATTEVAEV 135  
 S A+KP Y+PFM E+ LSKI++ GVMAL++YG D+K+Y+L QKSS+K+ A++++VA++  
 Sbjet: 61 SQAQKPSYIPFMCEEDVLSKIKETIGVQVQVGNALISYGDIDKVTYHLQKSSAKISATTEVAEV 120

50 Query: 136 RMTPIFGSDPFDKDPIDQKQVITYLEKSLYSLFPPKDGGLKPFVEMGNPFRVIGVFESK 195  
 ++ ++G F + F +Q+QV YLEKSLY++LFPK DG+GK+VRV GNPF+VIGVFES  
 Sbjet: 121 KQRLLEGEFGDSEAFKNCQVAYLEKSLYDYLFPKDGGLKPFVEMGNPFRVIGVFESK 180

55 Query: 196 EQSGLTSGTEKIAIYPLHCYVINGVNDATPEITITQYTRADDLPKVAKRVSMDMNQTPK 255  
 EQSGLTSG+EK+AYIDL QW+ I ++ +PE+T+QT++ADDLK VAK+VSD LKQ +P+  
 Sbjet: 181 EQSGLTSGSEKVAIYPLQWHRIFDTINVSPEVTQTHKADLLKVAKVSVDYNDQNPQ 240

60 Query: 256 SDYMPGVNMLKQFERQDLNLSNIFVLLAGIASISLIVGGIGWNIMLVSVTERTREIGI 315  
 SDYMPGVN+NL+RERQDLNLSNIFVLLAGIASISLIVGGIGWNIMLVSVTERTREIGI  
 Sbjet: 241 SDYMPGVNMLKQFERQDLNLSNIFVLLAGIASISLIVGGIGWNIMLVSVTERTREIGI 300

65 Query: 316 KKALGARRKLI 326  
 KKALGARRK++  
 Sbjet: 301 KKALGARRKIL 311

65 An alignment of the GAS and GBS proteins is shown below:

-525-

Identities = 79/386 (20%), Positives = 170/386 (43%), Gaps = 38/386 (9%)

Query: 5 ALAYITRKQNTLIIIFAILTIVLSCLYSCLTIMKSSNB-IRKALYESSNSSISITKXGD 63  
 A++ I K R+++ + I + + + +I++ + B ++ L SN++I+I  
 5 Sbjct: 7 ALSSLSGHOMBSILTMGLIIIGALIAIFSIIBGNTRMTKRLIGSNMNTINIV---- 61

Query: 64 YFNINQFMNIEKIKEVEEKIPQYDGLAKLKDVKVSGEISNRDLSDFNKVVLSRATS 123  
 FN K ++ K F AR D E+ +++ KN +  
 10 Sbjct: 62 -FN-----KKSSIDPK-FPOKSNARKPDYLPFMARELSKIQQVGKVNALLISYGD 111

Query: 124 NTKRNLFFSSGVSPFKBGKNIENDKNSILVHEEPAQKNKLGLDEIDLELDE----- 178  
 + +L S KN+ E + + + +F+ ++ + I LE E  
 10 Sbjct: 112 DKVYHLGQKSSAKISAITKNVAEVRMTFIRGSDFSKDFIDQKVILEKSLYESLPFK 171

Query: 179 -----KSGKIKSHKFKIIGFSGKKQETTYGLSSDFSENMFVDTSTSQEILANKSNENRI 233  
 K ++ + F++IG+F K+Q +GL+S +E + ++ I +  
 15 Sbjct: 172 DDGLGKFVEMGNFRFRVGVFBSKEQ---SGLTSG-TBKIAIYPLHQWYINGVDEATPE 227

Query: 234 ANKILMYSQSLSESTELAKLKDPKDKSKYSIKDN-KAPESLESVSGIKHLIK--IN 290  
 + I+ ++ + + I KS Y N K FE I++++ ++ I  
 20 Sbjct: 228 ITIQTIRADDLKFAKRVKSDMLNQTIPKSDYMFVGNLKEPERQLDNLNKSNFVLLAGI 287

Query: 291 TYSIMLGIVVLILLILNLRERIVEIGIFLSIGTTKIQIROPILIPIS----IPSI 346  
 + S++GGI V++++L+ + ER EIGI +Q + I++Q+ E + ++ + +  
 25 Sbjct: 289 SISLIVSGIGVNNILVLS-VTERTRIGIKNALGARRKLILKQFLIEAVILTLGGVIGV 346

Query: 347 ISSLFIQNLILVIVVEGPIINSNMI 372  
 IS + G ++ + + +I S S++  
 30 Sbjct: 347 ISGVSGLIITRSELYPYILSLFSVV 372

A related GBS gene <SEQ ID 8571> and protein <SEQ ID 8572> were also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Crend: 10  
 McG: Discrim Score: 5.59  
 35 QvH: Signal Score (-7.5): -5.97  
 Possible site: 56  
 >>> Seems to have an uncleavable N-term signal seq  
 ALOM program count: 4 value: -11.52 threshold: 0.0  
 40 INTEGRAL Likelihood = -11.52 Transmembrane 294 - 310 ( 285 - 312)  
 INTEGRAL Likelihood = -10.67 Transmembrane 396 - 412 ( 385 - 417)  
 INTEGRAL Likelihood = -8.76 Transmembrane 17 - 33 ( 14 - 38)  
 INTEGRAL Likelihood = -4.14 Transmembrane 335 - 351 ( 333 - 357)  
 PERIPHERAL Likelihood = 4.51 315  
 modified ALCM score: 2.80  
 45 \*\*\* Reasoning Step: 3  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.5607(Affirmative) < succ>  
 50 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

38.7/67.3% over 421aa Streptococcus  
 55 pneumoniac  
 GP[5712667] Vexpl Insert characterized  
 ORF00815(304 - 1575 of 1875)  
 60 GP[5712667]gb|AAD47592.1|AF140784\_1|AF140784(4 - 425 of 425) Vexpl {Streptococcus  
 pneumoniac}  
 %Match = 25.0  
 %Identity = 38.7 %Similarity = 67.2  
 Matches = 164 Mismatches = 136 Conservative Sub.s = 121  
 65





-527-

**Example 420**

A DNA sequence (GBSx0456) was identified in *S.agalactiae* <SEQ ID 1363> which encodes the amino acid sequence <SEQ ID 1364>. Analysis of this protein sequence reveals the following:

```

Possible site: 26
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
10 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 421**

A DNA sequence (GBSx0457) was identified in *S.agalactiae* <SEQ ID 1365> which encodes the amino acid sequence <SEQ ID 1366>. Analysis of this protein sequence reveals the following:

```

Possible site: 17
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
25 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAA74029 GB:U30715 ORFB [Bacillus anthracis]
Identities = 33/76 (43%), Positives = 44/76 (57%), Gaps = 1/76 (1%)

Query: 11 IRRVSHACTKAGDRFYENILAREPTATAHQKACTDVTYLOVGLGAKAYLSAIDLYNG 70
 ++R R EN+LRR F A N+KW TD+TYL +G YL +I DLYN
Sbjct: 86 VXRKRRTWINGESRIVENLLRRNFPQAKPKHGVTDITYLPFGT-EMLYLLSMDLYN 144

Query: 71 SIAYEISHNHILHL 86
 IIAYEIS+ ++ L+
Sbjct: 145 EIIAYEISNRQDVTLV 160

```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 422**

A DNA sequence (GBSx0458) was identified in *S.agalactiae* <SEQ ID 1367> which encodes the amino acid sequence <SEQ ID 1368>. Analysis of this protein sequence reveals the following:

```

Possible site: 27
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -0.69 Transmembrane 10 - 26 (10 - 26)

----- Final Results -----
 bacterial membrane --- Certainty=0.1277 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
50 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

-528-

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### 5 Example 423

A DNA sequence (GBSx0459) was identified in *S.agalactiae* <SEQ ID 1369> which encodes the amino acid sequence <SEQ ID 1370>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4170(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AA56999 GB:U09558 ORF4, putative Helix-Turn-Helix motif from
 amino acid 21 through 42 and from amino acid 78 through
 99 [Lactobacillus johnsonii]
Identities = 28/116 (24%), Positives = 59/116 (50%), Gaps = 6/116 (5%)

Query: 3 YSTLAKEQGVQYLDGKGLRDIKWDYDISSRVLQKWIKRYSGEDLKATSRGYSRMKQ 62
 YST K + V YL+ + S++ + K Y+I +++++W+ + + L A S +++
Sbjct: 4 YSTELKIEIVSKYLNHEDGKGLAKQYNTHM-FLIRWVDK-AKQGLAALSVEHTKITTY 61

Query: 63 GRQATFEERVEIVMTYIAHGKDYQAATEKFGVSYQQIYSWVRLEKNGSQGLVDRR 118
 + ++ +V Y + H KP +S Q+Y+W +K + G GL+ ++
Sbjct: 62 SS----DFKLNVRVRYLTHSIGSVKVAKFNISDSQYVNAKKFNEEGYAGLLPQK 113
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 424

A DNA sequence (GBSx0460) was identified in *S.agalactiae* <SEQ ID 1371> which encodes the amino acid sequence <SEQ ID 1372>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -0.69 Transmembrane 2 - 18 (2 - 19)
```

```
----- Final Results -----
bacterial membrane --- Certainty=0.1277(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 425**

A DNA sequence (GBSx0461) was identified in *S.agalactiae* <SEQ ID 1373> which encodes the amino acid sequence <SEQ ID 1374>. This protein is predicted to be integrase (phage-relatedpr). Analysis of this protein sequence reveals the following:

```

5 Possible site: 28
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
10 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

15 >GP:AACT79517 GB:U88974 CRF1 [Streptococcus thermophilus temperate
 bacteriophage 01205]
 Identities = 104/172 (60%), Positives = 127/172 (73%), Gaps = 11/172 (6%)

Query: 10 QHCSYAALYLIAGTGRFAECLGLTVNDIDYINKYLSINKTWDPYHPNQRYLPTNKSSIR 69
 ++ SYAALY+I+KTG+RFAECLGLTV+DI LG+NKTWDPY N ++PTK KSSIR
20 Sbjct: 186 EYASYAALYLIISKTGIRFAECLGLTVDDIKRDTGMLSVNKTWDPYKNTGFMPTKTSIR 245

Query: 70 NIPIDNDITLFFHEPTTNKNDLPDKLSNNAVNKTKIRKITGREVKVHSLRHTFASY---- 125
 IP+D++ + F+ + + RL LGNNNAVNKT+RKI GREVKVHSLRHT+ASY
25 Sbjct: 246 RIPLDDEFINFDQLPPTDDGRLPLSLNNAVNKTLRKIVGREVKVHSLRHTYASYLIAH 305

Query: 126 ---LISISQVLDEHNLNITLEVYAHQLQBQKDRNDKLNQRLGRINGKIALN 174
 LIS+SQVL HENLNITLEVYAHQLQBQK RND+ + ++W K N
30 Sbjct: 306 DIDLISVSQVLGEHNLNITLEVYAHQLQBQKSRNDE----KIKQMWTQSQN 353

```

30 There is also homology to SEQ ID 578

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 426**

35 A DNA sequence (GBSx0462) was identified in *S.agalactiae* <SEQ ID 1375> which encodes the amino acid sequence <SEQ ID 1376>. Analysis of this protein sequence reveals the following:

```

 Possible site: 22
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
40 bacterial cytoplasm --- Certainty=0.3206 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

45 The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 1328.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 427**

50 A DNA sequence (GBSx0463) was identified in *S.agalactiae* <SEQ ID 1377> which encodes the amino acid sequence <SEQ ID 1378>. Analysis of this protein sequence reveals the following:

```

 possible site: 45

```

-530-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5           bacterial cytoplasm --- Certainty=0.6542 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

10 >GP:CB52541 GB:AJ131519 hypothetical protein [Lactobacillus  
           bacteriophage phi adh]  
      Identities = 24/55 (43%), Positives = 36/55 (64%)  
  
      Query: 12 MDKELTPQSKANKKWAANNRRRTYLSKRSTARSFINNNATKEDLLEKQLIESK 66  
           M K   + KANKK E N+   + Y++KRSTA+SFI   AT+EDL +++ + +  
 15       Sbjct: 1 MAKITEPAKANKKNDKKNKARKLYINKRSTANSFILNATEDLANIEYVAER 55

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## 20 Example 428

A DNA sequence (GBSx0464) was identified in *S.galactiae* <SEQ ID 1379> which encodes the amino acid sequence <SEQ ID 1380>. Analysis of this protein sequence reveals the following:

Possible site: 41  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

25           bacterial cytoplasm --- Certainty=0.4417 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

30 The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 1332.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## 35 Example 429

A DNA sequence (GBSx0465) was identified in *S.galactiae* <SEQ ID 1381> which encodes the amino acid sequence <SEQ ID 1382>. Analysis of this protein sequence reveals the following:

Possible site: 28  
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

40           bacterial outside --- Certainty=0.3000 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

45 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 430**

A DNA sequence (GBSx0466) was identified in *S.agalactiae* <SEQ ID 1383> which encodes the amino acid sequence <SEQ ID 1384>. Analysis of this protein sequence reveals the following:

```

5 Possible site: 47
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -4.30 Transmembrane 205 - 221 (202 - 223)
 INTEGRAL Likelihood = -3.56 Transmembrane 296 - 312 (294 - 312)

 ----- Final Results -----
10 bacterial membrane --- Certainty=0.2720 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9663> which encodes amino acid sequence <SEQ ID 9664> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8573> and protein <SEQ ID 8574> were also identified. Analysis of this protein sequence reveals the following:

```

20 Lipop: Possible site: -1 Crend: 8
 MoG: Discrim Score: -0.80
 GvH: Signal Score (-7.5): -4.03
 Possible site: 47
 >>> Seems to have no N-terminal signal sequence
25 ALOM program count: 2 value: -4.30 threshold: 0.0
 INTEGRAL Likelihood = -4.30 Transmembrane 205 - 221 (202 - 223)
 INTEGRAL Likelihood = -3.56 Transmembrane 296 - 312 (294 - 312)
 PERIPHERAL Likelihood = 2.97 20
 modified ALOM score: 1.36

30 *** Reasoning Step: 3

 ----- Final Results -----
35 bacterial membrane --- Certainty=0.2720 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

SEQ ID 8574 (GBS366) was expressed in *E.coli* as a GST-fusion product. The purified fusion protein (Figure 215, lane 5) was used to immunise mice. The resulting antiserum was used for FACS (Figure 281), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 431**

A DNA sequence (GBSx0467) was identified in *S.agalactiae* <SEQ ID 1385> which encodes the amino acid sequence <SEQ ID 1386>. This protein is predicted to be N-acetylmuramoyl-L-alanine amidase. Analysis of this protein sequence reveals the following:

```

50 Possible site: 31
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1471 (Affirmative) < succ>

```

-532-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 5 A related GBS nucleic acid sequence <SEQ ID 8575> which encodes amino acid sequence <SEQ ID 8576> was also identified. This has an RGD motif at residues 81-83.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB07986 GB:Z93946 N-acetylmuramoyl-L-alanine amidase  
 [bacteriophage Dp-1]  
 Identities = 99/140 (70%), Positives = 120/140 (85%)  
 10 Query: 10 MVINIEQALAWMASRKKVTVYSMDYRNGPSSYDCSSSYFALRSAGASDNGWAVNTEYEH 69  
 M ++IE+ +AVM +RKG+V+YEMD+R+GF SYDCSSS+Y+ALRSAGAS GWAIVNTEY H  
 Sbjct: 1 MGVDIEKGVAVMCKARKGRVSYSMDFRDGPFUSYDCSSSMYALRSAGASGAWAVNTEIMH 60  
 15 Query: 70 DMLIRNGTVLIARNTNWAQRGDIPIFNGRGSAGAGFHTGMFUDPNIIHCPKGYTSIT 129  
 WLI+NGY LI+EM W+A+RGDIFNG+GASAGR GHTGMP+D DNIICHY Y+ I+  
 Sbjct: 61 AMLIRNGTVLILSENAFMDAKRGDIPFNGRGSAGAGGHTGMFIDSDNIIHCHYVDGLS 120  
 20 Query: 130 VNRHDEINGVNGQPPYVYR 149  
 VN+HDE W Y GQPY Y YR  
 Sbjct: 121 VNDHDERWYTAGQPPYVYR 140

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1387> which encodes the amino acid sequence <SEQ ID 1388>. Analysis of this protein sequence reveals the following:

25 Possible site: 26  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -1.06 Transmembrane 79 - 95 ( 77 - 95)  
 30 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1426(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 35 An alignment of the GAS and GBS proteins is shown below:

Identities = 56/91 (61%), Positives = 68/91 (74%)  
 Query: 158 KVINQSVVSKFKELDVNTPLNSNMFPYEATISDYVVEKPDVNSTOKELLVAGTRVR 217  
 K+D F ++LD NT L NSN+PYEAT+ DYVVEKPD+ +S DKE + AGTRVR  
 40 Sbjct: 354 KIDKFPQSQLTFNQKLTNTWKLKNSNFPVYEATLRIDYVVEKPNASADKEPIKAGTRVR 413  
 Query: 218 VYEKVGWARIQAPQSNQWVEDAYLIDATM 248  
 VYEK GW+RI A QS+QWVED YL+AT +  
 45 Sbjct: 414 VYEKVGWSRLNASQSQWVEDKYLGNATQV 444

SEQ ID 8576 (GBS301) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 9; MW 30kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 3; MW 55kDa).

- 50 The GBS301-GST fusion product was purified (Figure 205, lane 4) and used to immunise mice. The resulting antiserum was used for FACS (Figure 300), which confirmed that the protein is immunoreactive on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 432**

A DNA sequence (GBSx0468) was identified in *S.agalactiae* <SEQ ID 1389> which encodes the amino acid sequence <SEQ ID 1390>. Analysis of this protein sequence reveals the following:

```

5 Possible site: 53
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -6.53 Transmembrane 8 - 24 (3 - 25)

 ----- Final Results -----
10 bacterial membrane --- Certainty=0.3612(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 30 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 433**

A DNA sequence (GBSx0469) was identified in *S.agalactiae* <SEQ ID 1391> which encodes the amino acid sequence <SEQ ID 1392>. Analysis of this protein sequence reveals the following:

```

20 Possible site: 34
 >>> Seems to have a cleavable N-term signal seq.

 ----- Final Results -----
25 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 30 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 434**

A DNA sequence (GBSx0470) was identified in *S.agalactiae* <SEQ ID 1393> which encodes the amino acid sequence <SEQ ID 1394>. Analysis of this protein sequence reveals the following:

```

35 Possible site: 36
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
40 bacterial cytoplasm --- Certainty=0.0120(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 435**

A DNA sequence (GBSx0471) was identified in *S.agalactiae* <SEQ ID 1395> which encodes the amino acid sequence <SEQ ID 1396>. Analysis of this protein sequence reveals the following:

```

Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4757 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9661> which encodes amino acid sequence <SEQ ID 9662> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 436**

A DNA sequence (GBSx0472) was identified in *S.agalactiae* <SEQ ID 1397> which encodes the amino acid sequence <SEQ ID 1398>. This protein is predicted to be a minor structural protein. Analysis of this protein sequence reveals the following:

```

Possible site: 23
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -2.39 Transmembrane 349 - 365 (347 - 366)

----- Final Results -----
bacterial membrane --- Certainty=0.1956 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAF43531 GB:AP145054 CRF39 [Streptococcus thermophilus
bacteriophage 7201]
Identities = 212/666 (31%), Positives = 323/666 (47%), Gaps = 52/666 (7%)

Query: 10 WGNLNLLELSAMNKP---NIASNTSTVNVQVFL----RMSSYGYVISIGETRLKLTVD 61
W NN + W +I +NTS V ++ L + Y + E ++
Sbjct: 5 WNNDRGYRIRLWVDQVQDIQNNTSCVRLRLSLNLTTFATQYSCSAFVFNQGLNMS 64

Query: 62 GRAETINNVNPSINYQQRKLLPAKDYIVNHNISDGNKPLNIGAYYPIN--PSNYGEATNQ 119
G + N +I L + V H DG+ +F + A++ + +S NQ
Sbjct: 65 GSPVLGNNTQI-----LIDQITTVRHADDGSG-VPGVHAHFNSSGWSFGRLDIGN 117

Query: 120 SISLPEKINRLSVSSAISGVLGNAVFTITNKYSTPETHNLKYDFKSGTGTIATGVGTSLW 179
I+L I R S G +GN V I+I+R TH L+Y ++ G IA VGTSY W
Sbjct: 118 QITLTTPRGSSSVRVSDGFIQNQVDISIDRIGGATHTLYAWENKQKLIADNVGTSTYK 177

Query: 180 TIPPTFANLLPNELTGTGNLIVETMDGSAGIKETKYTLSTIPNTATPKLSSITLSIT 239
TIP FAN +PN +G G +V+T I TL+ ++ T KP + TL+DT
Sbjct: 178 TIPEDFANLIPNSTSGRGTIYVDYITNGFINLOTOSTTLTASV-ITNKLKPSFTGPTLIT 236

Query: 240 NTLTSSIVSG-NNFVRIISKVKVDPGSAIGNNGSTITTSNAEIVGKSNSIIGNGVDFDL 298
N + IV G +FV I+S VEV P A +G+TI Y AEIVG +NSI NG V ++
Sbjct: 237 NPTSRIVPGQTHFVSISSLVKVFNPGAQAKSGATIVGYARTVGNNSISNGGVLRV 296

```



Query: 299 DFFGSA--TIRATVTDKRLTSEPVDTKINVIDYFLPIVTSKVVRSSQNPDLQLVLPFV 356  
 T+R V DSRG+ S+ V+TK+ + YF P+ + +V RS+ DIL+ F  
 5 Sbjet: 297 SVNQDTENTLKRGRVQDSRGIWSDWVETKLTFLFYFSPAL--RFEVKRSOKKLDLITKREFA 355

Query: 357 KIAPLIVGGIQKQLQMSVSAFVNTGIYAVDGAATNTWMTISQMSGAPLMLGQITDKS 416  
 KIAP+ V GIQ+N +K++ S A + + VD+G A WS+IS+ + + LG +Y  
 Sbjet: 356 KIAPLSVNTGIQKQVNMKLTSTAKVGMNRFVVDNGQAGQVWSISSEFNASDAKLGNSYPAD 415

10 Query: 417 KSNLVKISVSDNLMGATPIIQPIASEFVLVTAPSGVARGKLVHSGITIDAKGDVYVDGTI 476  
 S+V + D S T + V ++ V++? GV GK R G +D GD I  
 Sbjet: 416 TSYVVIKLEDEFTS--TSFQATVPTDEVIMTYDRCQVGLIGKYRERGALDVAGD-----I 468

15 Query: 477 YCGDKAIQKPLALANGGSPRHDDTLNLSIQDTGFCYVPRGANRPAGRGYIVTVRIET 536  
 Y + IQQ L NNG ++ N+++D G Y +F A P + + + H +  
 Sbjet: 469 YANNSPIQQYQLTNNGGSPMTANA--NTIEDPQQYILFSA--PONPSQNGHILPHISS 524

Query: 537 -----ANYAQCFYDRNKTI-----FTRLLENGVNSGWSEYVKGD--SIQTGMITIG 583  
 A Q F+ + +R+++ W N E+ + D +L TGN G  
 20 Sbjet: 525 YGKSGMYKEAIQFWSNDGRLFSKHHSWSRI ID--HEPWKEPARNDWNLINTGWQPAQ 582

Query: 584 -NGFKYRKGGDDIDIMYFASNGLQRWSVGWMPGSLI--POELNFAITGNTLAPDKS IHL 640  
 +G YKR GD + + +NF G + + +P + PQ NF +TGN+ +XK ++  
 Sbjet: 583 VDGFSYKRVGDVLTIKNFTGTG--GDFLLASVPPETIFAPQSYMFFVTVGWSVMANKQTVN 641

25 Query: 641 QDNAGS 646  
 Q+N G  
 Sbjet: 642 QVNEGS 647

30 No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 1398 (GBS365) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 82 (lane 2; MW 102kDa).

GBS365-GST was purified as shown in Figure 216, lane 11.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 437

A DNA sequence (GBSx0473) was identified in *S.galactiae* <SEQ ID 1399> which encodes the amino acid sequence <SEQ ID 1400>. This protein is predicted to be a minor structural protein. Analysis of this protein sequence reveals the following:

40 Possible site: 59  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 45 bacterial cytoplasm --- Certainty=0.3481 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

50 >SP:AAC34413 GB:AF158600 putative minor structural protein  
 [Streptococcus thermophilus bacteriophage Sfil1]  
 Identities = 504/998 (50%), Positives = 675/998 (67%), Gaps = 56/998 (5%)

Query: 1 MLTHTGPDLEKPVLFPLNDKQKALNYFNHKKYRKOKTGSSVLEFSVYKDKLLGDSPLSHKY 60  
 +LTH +L+ V ++DN+KQ LN+FN KN R ++G+SV EFSV+KK + DS + Y  
 55 Sbjet: 2 LLTTHDNHNLQKVAVIDNKSQSTLNFNFDNDRSLESSTSVFESVFKKSLSKDSKVEIST 61

Query: 61 HVLNDQAFVSPVHKKGKVLNLNINKIDEDERKQIDCYCENTINLELMRYCNAYKATKAMSFE 120

-536-

LN++AFVSF HKKG L N+MKI+RDE+ I CYCERN+LELL EY AYKA+K M+F+  
 Sbjet: 62 KYLNREAFVSFHKKGKYLFPNVMKISBDSQIIRCYCNLSLELLLEYRGA YKASKMPMPFK 121

5 Query: 121 EYLVPQDILSGALITVOTNEVKKKLTLWSTQETKLARLGLSIAMNPDAIEFETKLNPN 180  
 EY + + + LT+G REV D+K TLEW QET LARL+S+A NFAEIRF+T+L N  
 Sbjet: 122 EYFDWGWGQPAKILGLVNSVSDQKRTLWESQETTLARLISLARNFARLISFDTRKLPK 181

Query: 181 HTFKQLIINTIKEYEBGKSYGVDRDKTQVILARYQKNISGIRKTVKQIYNARIPYGGK- 239  
 + ++N+YK Y+ GK+ GV R ++DVL+Y KNI+GI+++VDK QIYN I PFG+K  
 10 Sbjet: 182 SQLEDFVIANVYKAYD-GKMGCGRRSSDVLACGKNGNIGKRSVDKTIQYNNIMITPYGKRS 240

Query: 240 -TVRGERVSNFVTRKVTYVGSNRT---YLGEDLKYYGHTIKKANVQALINAYOYNIL 295  
 T + + IS+PVT + V S R Y GDL Y GHT+ + VQ I N VQ N+L  
 Sbjet: 241 DTKKTKRISDPTVITQNPVVPFSARVEKRYAGSDGLTYAGHTLSASLVQITPMLCVQRNLL 300

15 Query: 296 PGGIITQLYLSFWDSTVGRKRRNMGSGGAGTRPSGVKVTTCMARFANEGGTMYTA 355  
 PGG+I+QLYLSFWG S V +RNNW+GH+GGQTRPSGV VTG RPA+EGSTMYTA  
 Sbjet: 301 PGGVISQLYLSFWGSSNVARRRNMGSGHTGGAPTRPSGVVVTGSPRPAEGSTMYTA 360

20 Query: 356 SVDDFLKDYTYLLAKQG-----INNVGKGNIAIDYTKLFRAGGAKDYAARGYSYTHL 410  
 SVDDF+KDYTYLLA Q +Y V GK+NI +YTKLFR CGA YDAAGY Y L  
 Sbjet: 361 SVDDFMKDYTYLLAQDTSGRKMGYGVKGNIEYTKGLFRIGGALDYAAGYNYHYIL 420

25 Query: 411 MTNRNGINKVTRNLTATIDKLWQTPUKPTANVARRATKTIQ-----INZATKLG 464  
 M +IRNGIN+ GHTL +D LN+ P IT N ++ T+T+A +NE LKG  
 Sbjet: 421 MRDIRNGINRNGNILLDKLDDLWQPDWQITQPN--KQVTKTKADEIAVLNMGQLKG 478

Query: 465 KRIGSGQCYALSGWAKKLDGAWIDSSGIGRIGCGMAALIGTYNMGYGMVKVKS 524  
 RR+G+GQCYAL+ NY+ KL G + + + G IG GMAA ICTDY W +GM V +  
 30 Sbjet: 479 RRVGNGQCYALAAWYMKLGGPGLGAGVTGSGVGIGMAAAKIGTYDWRFGWSVRP 538

Query: 525 PNAGNLKAGGIYNVNRNGAPFYTTWGHTGLIKSVKTRVTLGQNFVGRMYVNSYD 584  
 + LK G I N+A+ T+ WGH II S + + VTVLGN+ GR YVV+NEY  
 Sbjet: 539 TSDVQLKGAIAWIAKNSY-LSTSVMGRVSIILNNGSTVTVLQNYAGRGVYVQNSYP 597

35 Query: 585 INSEFASGLQTCYPRELAQMSVMGATQQVSGGTISYSEVVQSAQTESYREQLIYID 644  
 +++ ++T+CYP E+ +G +V G T + ++ E+ + E + ID  
 Sbjet: 598 ASAYLGAVELTCLPELKEGKTVEGRITVSTPNRVEQVKEIPIDVETTRSTALATID 657

40 Query: 645 NSIYKWKDENGKVEYLLKMFLYAPLGRDRYPBVLTCNETRDNWIKRDMVEVTDSEVL 704  
 + +EW+ENG+VE+YL+NG LYAP+S++ YP8+LTG E DMNWKDM2++TDS+VL  
 Sbjet: 658 SKRKGWRNENGQVEFYLENGSLYAPISKELYP8ILTGKENGDNWIKRDMIDTDESVL 717

Query: 705 MSTGLKDLKAHAYPAITYEDGVVDLELGDVVRIGQGYEPPLILTARVSEORISITNPS 764  
 +ST L+L+ YPAITYEDVG++DL+GD V+IQD G+ P L+L ARV EQ+IS TNP  
 45 Sbjet: 718 ISTALRNLKFCYPAITYEDVGFLLDIDGDTVKIQD GSPMLMLARVSEBQSLTIPN 777

Query: 765 SSKTKFSNFVEKESQLASDLISDMRLYDESIPYEIKLATENGVAFNGTGRSVLTSIQ 824  
 NKT F+NF +++++ L+S M +L +E+IPYE+KL+T NG FEN TG+SVL +L+  
 50 Sbjet: 778 ENKTVFANPTQLNKVSDSLSRMTKLAEALPYELKLTSTNGTTFKNGTGSVLKATL 837

Query: 825 KNGKDYEAIFYKNGDSLIDIGPSLIVKASDPNRVLNITVEAYLNRELVASTQISFIDTE 884  
 KNG+ Y+ ++F+KNGDS+I G L+VK +DF + L +TVEAYL++ELVAS +I+PTD  
 Sbjet: 838 KNGEVPQIPIFFKNGDSLIIOTGNQLVWPKTDFENTLQVTVEAYLDELVASAETITDVS 897

55 Query: 885 DGADGKDGAPGPGQPGVNGIQGPKGDQGIQGPAGDGKATYTHAYALDENGSTGFSVS 944  
 D3 QGPKGD G+ L S+G  
 Sbjet: 898 DCK-----CGPKGDGVS-----PINLISSNGYQFK 925

60 Query: 945 DNVGKTYI--GMVVDENIIDSNDPK-KYKGNLKGADG 979  
 +N+ T +Y D+ ID + + Y W+ + ADG  
 Sbjet: 926 NNIINTTFTAKLYQDNKIKDKGTRYAYLMSKV-NADG 962

A related DNA sequence was identified in *Spyogenes* <SEQ ID 1401> which encodes the amino acid  
 65 sequence <SEQ ID 1402>. Analysis of this protein sequence reveals the following:

Possible site: 37

-537-

>>> Seems to have a cleavable N-term signal seq.  
 INTEGRAL Likelihood = -3.56 Transmembrane 325 - 341 ( 323 - 343)

----- Final Results -----  
 bacterial membrane --- Certainty=0.2423(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 23/55 (41%), Positives = 27/55 (48%)

Query: 886 GADGKDGAPGPGPGVINGLQGPKGQGIQGPAGADGKATYTHIAYALDENSTG 940  
 G GKDGPAG G PG G +G +G+ G QGP G G+ T G G

Sbjct: 181 GEAGKDGAPKQDAPGEGKGEKQDRGETGAQGPVGPQGERGSTGAQGPAGPQGEAG 235  
 Identities = 48/151 (31%), Positives = 58/151 (37%), Gaps = 19/151 (12%)

Query: 852 KASDFNHFVNITVEAYLNE--ELVASTQISFDTEDGADGKDGAPGPGPGVINGLQGP 909  
 K DF D E L E +L+ + I + G G G GPQG G G QGPK

Sbjct: 82 KEEDPQKELKDFTEKRLKEILDILGKSGIK---GDRGETGPAAGPAGGKGTGERGAQGP 138

Query: 910 GD---QGIQGPAGADGKATYTHIAYALDENSTGFS---VSDMWKTYIGMYVDNIID 962  
 GD QGIQ AG G+ E G G + GK D

Sbjct: 139 DGRGEQGIQKAGBKEGKEGKDRGETGERGERGEAGIQQPQGEAGK-----DGAPGK 191

Query: 963 ENDPKCKYKNLIKADGARGIQQPAGADGKT 993  
 P + +G GA+G GP G G+T

Sbjct: 192 DGAPGEGKEGKDRGETGAQGPVGPQGERGET 222  
 Identities = 25/50 (50%), Positives = 29/50 (58%), Gaps = 9/50 (18%)

Query: 884 KDGADGKDGAPGPGPGVINGL-----QGPKGQGIQGPAGADGKA 924  
 +DGA GHDGAPG +G G G QG K+ G QGPAG G+A

Sbjct: 185 KDGAGHDGAPGKGEKQDRGETGAQGPVGPQGERGSTGAQGPAGPQGEA 234

SEQ ID 1400 was expressed in four different forms. SDS-PAGE analysis of total cell extract is shown in Figure 122 (GBS105dN – lane 5 & 7; MW 102kDa), Figure 122 (GBS105dC – lane 8-10; MW 81kDa), Figure 179 (GBS105d – lane 8; MW 102kDa) and in Figure 181 (GBS105C – lane 2; MW 56kDa). GBS105dN-His was purified as shown in Figure 232 (lanes 9 & 10). GBS105dC-His was purified as shown in Figure 233 (lanes 3 & 4).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 438

A DNA sequence (GBSx0474) was identified in *S. agalactiae* <SEQ ID 1403> which encodes the amino acid sequence <SEQ ID 1404>. This protein is predicted to be a minor structural protein. Analysis of this protein sequence reveals the following:

possible site: 32  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2502(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AA34412 GB:AF158600 putative minor structural protein  
 [Streptococcus thermophilus bacteriophage Sfil1]  
 Identities = 163/433 (37%), Positives = 244/433 (55%), Gaps = 21/433 (4%)

-538-

Query: 80 LSSKKPKMLPSPHSPQRYYLAVQGLNFKPKIKMNGPGRIT--FIVADAYANSTSYRIK 137  
 L +KK L P RYLA+ G+++ K I + + R T F+V D AHST+Y+R+  
 5 Sbjct: 93 LHTKGA/VK/LFTEPERRYALVAGEVSLGIS-DWYDEATIEPLVPGGVARSITYTKRV 151

Query: 138 DYTQDGNKMTFKIKNGTAPAPPIPRIKHLGNGVIGITNGTGAFAVGSPEHEOTIVHR 197  
 DY + KM F I N G+ A+PI +K ENG Y G + + AF G+ EE DG I+ +  
 Sbjct: 152 DYQKDGKMIPLSDNGSTDAYPLITLWANAENGTYGLVSDKFAEAGNIEADGKIISK 211

10 Query: 198 NETLPDY-SKALQAAL-EGAPVAKINIMPPTFDSLKMKRLDNLGSGKGEYVAIGAR 255  
 E L+D+ I QA +GR NV N + + + N+ G IG +  
 Sbjct: 212 AEVLVDYFRDRIQAPAKGAKNVGIVTIGDLHOT---LSIQNVWRPH-----IGLK 261

15 Query: 256 GTTPGYGE-HVGRTRFIINPDNGEY-TLASHLWNRQIFITATQDQXGLKLCVTGENDE 313  
 + + T I PDS+G LNE+WN+QIP A+ Q GFLKL V+ +  
 Sbjct: 262 NPNANINQLQASITLIDIPDSSGNVGNALNEYIWNQIPWAGSISQYGLKLVSDADGN 321

20 Query: 314 FLYGIETYKANGFETENIFFALDDGVWRFIKQFEQA-DRYNHNPPSMRRAVEIP 372  
 FLYG+ET+KR G E+EN A D G+RF KQ+ F A + HNP+ R + +I  
 Sbjct: 322 FLYGVETFRSLGLESETNALADSGY-GFRFLKQWFLATYEDHNPFPNEPROMS-DIK 379

25 Query: 373 REEDKFRIFYNGAHNVHVTPSLKQKSRKIHLMGTCSDSSKYINLFEKVPEKMGVS 432  
 RE+DK Y+ G + + T+P +KQKS KIH L + S + + + F + + + K + +  
 Sbjct: 380 REDDKVTFVWQTYNTFTPIEKGKSAKIHLLTINI-PSKSFVTHAYPQDLLYIKINNA 438

Query: 433 HYNIVNKKYQPDDEVIIINFENITVTKIDIDISQDVLGSKMSIPPCSESLVHLSSWA 492  
 + + I N-Y G + IIN E+DT+ + + + + V GS IPPCES+ V S W  
 Sbjct: 439 FFEDIPIRYIQSNLIINSRDTLTINNLMLDEIVGSLNFPVTPGSSQIVVQSPWAK 498

30 Query: 493 ALPDISIDPEERY 505  
 P ++I+FEER+  
 Sbjct: 499 KKPSTVIEPEERW 511

No corresponding DNA sequence was identified in *S. pyogenes*.

35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 439

A DNA sequence (GBSx0475) was identified in *S. galactiae* <SEQ ID 1405> which encodes the amino acid sequence <SEQ ID 1406>. This protein is predicted to be PblA. Analysis of this protein sequence reveals the following:

Possible site: 57

>>> Seems to have no N-terminal signal sequence

45 INTEGRAL Likelihood = -7.11 Transmembrane 427 - 443 ( 424 - 445)  
 INTEGRAL Likelihood = -4.99 Transmembrane 449 - 465 ( 448 - 469)  
 INTEGRAL Likelihood = -2.71 Transmembrane 41 - 57 ( 38 - 57)  
 INTEGRAL Likelihood = -0.37 Transmembrane 361 - 377 ( 361 - 377)  
 INTEGRAL Likelihood = -0.22 Transmembrane 324 - 340 ( 324 - 340)

----- Final Results -----

50 bacterial membrane --- Certainty=0.3845 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

55 >GP:AGI18638 GB:AY007505 PblA [Streptococcus mitis]  
 Identities = 233/401 (58%), Positives = 296/401 (73%), Gaps = 17/401 (4%)

60 Query: 1 MATNLQAYVQIMPAKGLSGISIKTLDPASSAGSSAGSLGKGLIGILGSVIAAAKIG 60  
 MAT + QAYVQ+PSA+GI+G I L+PSAS+AG SAG LG L+G++ VIAAA IG

-539-

Sbjct: 1 MATEIAQAYVQLIPSRAGITGKIQSILNPFASAAQSGAGSLGSSLVGMVTKVIAAAGIG 60

Query: 61 EMVTKAIISSISBGAALQQSLGGVETLPKSNANLVKKYADABKTTGLSANAYMRSVTGFP 120  
KA S++ISBGAALQQSLGG+ETLFP +A+ VK YA+EAYKTTGLSANAYMRS+VTGFP

5 Sbjct: 61 ----KAFSAALISBGAALQQSLGGIETLPKGSADKRVGYANEAYKTTGLSANAYMEVVTGFP 116

Query: 121 SASLLQSLGGDTAKAAKVAHMMIDMADNSNMKMTSMESIQAQYGFARQNTYMLNLKIL 180  
SASLLQSLGGDT KAA+ AHMMIDM+DN+NMKMTSMESIQAQYGFARQNTYMLNLKIL

10 Sbjct: 117 SASLLQSLGGDTYKAAETAHMMIDMSIRANKMTSMESIQAQYGFARQNTYMLNLKIL 176

Query: 181 GYGGTQREMRLLSDAQKLTGKKYDISMLSDVYEAHAIQKIGITGTTAKEMATFTFGS 240  
GYGGT++EM+RL+DA+KLGT KYDI+NLSDVY AHAIC + ITGITAKESA+TF+GSS

15 Sbjct: 177 GYGGTQREMRLLSDAQKLTGKKYDISMLSDVYEAHAIQKIGITGTTAKESAATFTFGS 236

Query: 241 FEMKAAASNNLKGKMLGEDIKPSLKAALPDTTNSFVLMNFIMLNVKFGPSVISLTF 300  
FEMKAA++N+LKG+ALG+I PSL AL TIS F+ +NF+PM+ NVF G G V++ S

20 Sbjct: 237 FEMKAAASNNLKGKMLGEDIKPSLKAALPDTTNSFVLMNFIMLNVKFGPSVISLTF 296

Query: 301 ELIPKIVGFMKTSGSPSLMSQSIFISFVNGFLTAYAPITVAGKIFTDFVFMVQSGIP 360  
++ ++ G S + +S + G + F + G + + + +I G

25 Sbjct: 297 QIASQLFG-----DAFGSAVFDQLSRITGIFETP--FDMIFGSLSKQNDIITNTI-G 346

Query: 361 LQAGATFLVNLIDGILANLPQATS---AVSVISSPISML 398  
+ AT ++H+ D L I I ++ V ++ F+ L

30 Sbjct: 347 FSEBAQTQVNIADNRVTPEINISGAIGDVGVIGVDFVGL 387  
Identities = 112/386 (29%), Positives = 172/386 (44%), Gaps = 18/386 (4%)

Query: 235 TPTCSFEMKAAKSNLKGKMLGEDIKPSLKA---LPTTNSFVLMNFIMLNVKFG 290  
TT+ E++KA ++ + L E IK + L T V+ PI H++

35 Sbjct: 580 TTYEAYVESLKAMENAVVTFPSDLWESKEAATAWTLITTAVMVVQFIDGPMININ 639

Query: 291 POSVISLTFSELIPKIVGFMKTSGSPSLMSQSIFISFVNGFLTAYAPITVAGKIFTDF 350  
++ + + + G + S+ I II V G A L++ + +

40 Sbjct: 640 ISBELTQVNECIKLIFEGAMEPI-KSIFLAILIIIDLVTGNFQGLGADLSLWEGIKNG 698

Query: 351 VSPFMQSIGLLQAGATFLVNLIDGILANLPQATSVAVSIFSPISMLQANYPALIKGF 410  
+S + I +++ G+ N ++ I + SM + I

45 Sbjct: 699 ISLWECIKTYFSGVVDVIGYATCVFENFNVLSITWEIFKTAASMA--MEMISVTS 755

Query: 411 ELSYLVQGIHARLPDIVITVGKL---IAILAGIASNLKPKVLALGVQLLITFVKQILSV 467  
+++ L+Q3 + V + L I A A S L K L L G + V G +

50 Sbjct: 756 NLTLGLIQGAQNLANNFVFLSGLWENIKETASAAWSGL-KSLVLA--FINGLSGAQTA 812

Query: 468 IKGINEANNICK---LINAISIDLLSAGRAIMRGFLEGLDWDGIDQNFVGDAGWI 524  
+ + +++ K + N IK+IL AG+AI+ GFL GL+ W + NFGV TA WI

55 Sbjct: 813 WNNKQAVSDLVTKVTFNFGIKNINLWEGKAILNGFLGGLKSAMEGVNFVGGTANW 872

Query: 525 KDHKGPISYDRLLIPAGNAIMQGLHQGLVDKFKPVNHLVNGMAEETQSFGNPQLAFDM 584  
+DHKGPI YDR+LLIPAGNAIM L GL D FK VK V GM+ EI F L +

60 Sbjct: 873 RDHKGPIEYDRLLIPAGNAIMQGLHQGLVDKFKPVNHLVNGMAEETQSFGNPQLAFDM 932

Query: 585 DNNVNGFE-RIGTLNKNLSSQVTST 609  
+V E R+ + L Q + T

65 Sbjct: 933 TAVSTKNLEARLAMPASQLVQESKT 958

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1407> which encodes the amino acid sequence <SEQ ID 1408>. Analysis of this protein sequence reveals the following:

Possible site: 55  
>>> Seems to have no N-terminal signal sequence

|    |          |                    |               |                        |
|----|----------|--------------------|---------------|------------------------|
| 60 | INTEGRAL | Likelihood = -2.76 | Transmembrane | 458 - 474 ( 458 - 474) |
|    | INTEGRAL | Likelihood = -2.60 | Transmembrane | 483 - 499 ( 482 - 499) |
|    | INTEGRAL | Likelihood = -2.02 | Transmembrane | 429 - 445 ( 429 - 445) |
|    | INTEGRAL | Likelihood = -1.28 | Transmembrane | 397 - 413 ( 397 - 413) |
|    | INTEGRAL | Likelihood = -0.53 | Transmembrane | 739 - 755 ( 738 - 755) |
| 65 | INTEGRAL | Likelihood = -0.27 | Transmembrane | 356 - 372 ( 356 - 372) |

-540-

----- Final Results -----

bacterial membrane --- Certainty=0.2105(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5

The protein has homology with the following sequences in the databases:

>GP:ANB18717 GB:U38906 ORF42 [Bacteriophage rit]  
 Identities = 261/579 (45%), Positives = 359/579 (61%), Gaps = 63/579 (10%)

10

Query: 184 MKRLLSARKLPAAMGRKFDLSNYADVRAIHVLQDNMGTAGVAEAAKTFPSGSLAAMK 243  
 M+RLL+DA+KL G+K+D+SN++D+ +AIH +Q M I G A+EA TTFSSG +MK  
 Sbjct: 1 MQRLLTAQAQKLT---GQKYDISNPSDITQAIHAIQTEMDDITGTAKAEASTTFPSGSDSMK 57

15

Query: 244 SSFTIVVAGSLGDDIRPALGLAETISNLFQNFIRMVANIYFKGLPSAIGTFIGNAAPI 303  
 ++ +IV+ LSLG D++ L L TTS FLF NFIRMV NIFK LP AI TF+ AA  
 Sbjct: 58 AAMSIVLGNLSLGRDLGGPLNALVSTTSTFLKFNFIYVGNITFKALPGAISTFVSAAGKE 117

20

Query: 304 ITSQ-----FQGLMSSLG- ISIDLSPIT 325  
 ++BQ F L+SS+G IS + +  
 Sbjct: 118 LSSQLGNGISGSGSDFTAKFSSLSPLQGSFQITVSGLKFPDLSLSSIGFTSTQIMGVF 177

25

Query: 326 AKFAQIGQNLQ---PVFNELKTAFSQLPSFFTSIGSAVAVPIDIISGLARLFSQFEA 381  
 +K Q+ N+ FV + L AP QLPS F +I AV P+IDTI SG++RLDQSG +A  
 Sbjct: 178 SKLPQLFSNVISAVIPVISTLSVAVGQLPSLFEATSVAVQVMDITSSIGSLKPLSGIQA 237

30

Query: 382 LISAILPALQAGFSNFAAIVGPAISGVVDSFVGMWNAQPLISILSDALMPVQILGSFL 441  
 +ISA++PA+ G + I+GP+I +V+SVF MWN+ QFL +++++ ALMP FQ+LG+P+  
 Sbjct: 238 IISALVPAITIGITMMGLIGPSIDLWNSFVWMNSIQPLATVIAGALMPAFQVLGAFI 297

30

Query: 442 GGVVKGALMGVSHAFDAVKVAIQLVPIIDLLVQGLNFVQVFLSVIAEWIGVAIGFQNL 501  
 GGV+KGR++ +S FD ++V + +TPII ++ FVL+ +A+W+G AIG F N  
 Sbjct: 298 GGVILGAMIALSATPFDITRVVGFLLPIIAIAVLAKQEFAPVLATVQWGTAGFPAF 357

35

Query: 502 GTAGQQLSAFIKSAWNIQTAISTAGTITISTVIDYIKLAFSGAGSAVGVLNIPISAWMA 561  
 G AG L I SAN I++ IS+ + I +I+ K P+G QSA G L++ S AW  
 Sbjct: 358 GAAGTSLNGLTSAWNGIKSTISSVSVSGIGTINTAKAIFGIGSAGGALRAMISGANS 417

40

Query: 562 MCDAINVAKGISVINGIKSAPSSFS-----SLVSVGSASVNGVIDSISTING--- 611  
 + I+ G IS INGIKS FSS S++S V S + G+I SSTI G  
 Sbjct: 418 IRSIISVGSISGTINGIKSFFSSLGQSGNGLRSVMSVSGITVGISASSTISGIIID 477

45

Query: 612 -----LANIDISGAGAAIINGFLNKLKSAWGVKSVFSGIANMTAHHKGPISYDRVL 663  
 L NID+GAG A++GP+ GLKS W A K FV GIA+NI +HROPISYDR +  
 Sbjct: 478 GIKNIFNSLKNIDLAGAGRAVIDSGFVQGLKSTWEAGKFFVGQIADNIDKHGKPSYDRKI 537

45

Query: 664 LKPAKRAIMGSLATSLIDGFKVKNVSNVSGMADLDAIMI 702  
 L PAG+AIMGG N SL++ FK V+ NVSG+A + S +T  
 Sbjct: 538 LIPAGQAIMGSPNDISLIMENFKAVQKIVSGIAKQIQSAIT 576

50

An alignment of the GAS and GBS proteins is shown below:

Identities = 272/701 (38%), Positives = 371/701 (52%), Gaps = 91/701 (12%)

55

Query: 1 MATNLGQAYVQIMPAKGISGISTIKTLDPBASSAGSGAGSLGGKLGILIGSVIAAIIQ 60  
 MAT LQQAYVQIMPSA+GISG+ISK LDPBA SAG SAGSL+GG L+ ++G IAAQ IQ  
 Sbjct: 1 MATNLGQAYVQIMPSAGISGISTIKTLDPBASSAGSLGAGSLGGKLGILIGSVIAAIIQ 60

60

Query: 61 EMVTKAISSSISGAAIQSSIGVETLFPKSNANLVKKYADRAVYTTGLSNAVYMSVTFG 120  
 +M ISS++S GA LQGS CG++TL+K VK +A EAYK G+SN Y B  
 Sbjct: 61 KM-----SSAISAGADIQSSFGGIIITLYKGRATVAVKFAKAYKA-GISANTYAYSAVSM 115

65

Query: 121 SASLQSLGGDTAKAARVAMAMIMADSNMKTSMSSISQAYVQGFARQVYTNLDMKL 180  
 ASL QSLGGD AAK ANPA++IMADNS IMGT + SIQ AYQGFARQVYTNLDMKL+L  
 Sbjct: 116 GASLQSLGGDAVAAKAAHMAIMAMANSACMGTDTISTQMAVQGFARQVYTNLDMKL 175

65

Query: 181 GYGGTQSEMKRLLSDA+KL---TGKKYDISNLSVDVYKAHAIQKIGITITGTAKKRAITF 237  
 GYGGT+SEMKRLLSDA+KL GKK+D+SN +DV EAH +Q +GI G A+EA TTF

-541-

5  
 10  
 15  
 20  
 25  
 30

```

 Sbjct: 176 GYGITKEEMKRLLSDAEKLPAAMGKKFDLSNYADVVAIHLVQDNWGIAGVAAREAKTTF 235
 Query: 238 TGSFEAMKAASKNLLGHMALGEDIKPSLKA/P/TTSNVLNPFIMPLINVKRGSVISL 297
 +GS AMK++ N++ +LQ+DI+P+L+ L +TTSNF+ NPLFM+ N+FKG S I
 Sbjct: 236 SGBIAAMKSSPTNVAGLSGDDIRPALRGLAETTSNLFPGHFIPMVANIKGLPSAIGT 295

 Query: 298 TFSSLIPIKIV---GFM/TSGPSIMQSGSFIISFV-----KGFLTAY---PAPIAV 342
 P I G M + G S+ S I+ + + NG TA+ P+P T
 Sbjct: 296 FIGRAAPLIITSQFQELMSSIGISIDLSPTTAKFAQIQNLQPVFMGLKTAQSILPSFTS 355

 Query: 343 AGKIFTDFVSPVMQSIGL---LQAGATLVNMLDGLILANLPQIATSAYS-VISSFSIM 397
 G + ++ + L +A + +L + +N I +A+S V+ SP+ M
 Sbjct: 356 IGSAVAFVIDTIIISGLARLDSGSPALISAILPAIQAGFSNFAATVGPISGVDSFVGN 415

 Query: 398 LQANTPAI-----LKKGFELISYLVOGI-----IARLPDVIIT 430
 A P I L F+IL + G+ + + D+++
 Sbjct: 416 WNAUQPLISILSDAMPVFPQILGSFLGVVKGALMGVSFAFDVAVKVAQLVTPIDLLAQ 475

 Query: 431 ----VGKLIATLGAIASNLKFKVALGV---QLLITPFKGLISVIGKINETANNIGEKLN 484
 V ++++++A I + LG Q L F+K + I TA I + +
 Sbjct: 476 GLNFVQPVLSVIAEWIGVAGIMQGNLTAGQILSAFIKSAWNIQTALSTAGTIISTVID 535

 Query: 485 AIKSI-----DLISAGRAMRGFLRGLEDVMDIQNPVGDIA 521
 IK D + + + I+ + G+ + + + D + +
 Sbjct: 536 YIKLPSGAGSAGVGLKNIHPSLAWMAGDAINVAIGIISVINGIKSAPSPSSLVSSVG 595

 Query: 522 CMKDHKGPISYDRLLI----PAGNAIMQGLHQIOLVDFKPKVKNLVNGMAREIQSSGP 576
 + TS R L AG AIM G GL + VK+ V+G+A I G
 Sbjct: 596 EAVNGVIDSISSTIRGLANIDISGCAAIMQGFNLGLKSAWAVKSPVSGIANWIAEHKG 655

 Query: 577 NPQLAFMDTNNVNGFERIGTIAKNNLSQVTSSTWNTGNA 617
 +++D G +G LN +L + SG A
 Sbjct: 656 --PISYDRVLKDPAGKAIMGSLNLSLDIGPKVKENSVSGA 694

```

- 35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 440

A DNA sequence (GBSx0477) was identified in *S. agalactiae* <SEQ ID 1409> which encodes the amino acid sequence <SEQ ID 1410>. Analysis of this protein sequence reveals the following:

40  
 45

```

 Possible site: 16
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2565 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

 >GP:AMG18637 GB:AY007505 unknown [Streptococcus mitis]
 Identities = 64/119 (53%), Positives = 87/119 (72%), Gaps = 2/119 (1%)

 Query: 1 NLQNDDEALVCDLAETTHYIDYKLEPLKVAVFSLGLREESRINRVISGNRVSPERRILA 60
 N++ DEDAL+CDLAETTHY I+DY+QLP +VAVF+ GLR++SRI ++ +V FE +LA
 Sbjct: 1 NIQYDEADALCDLAETTHYIPIYRQLPADQVAVFAPGLRDSRIKLWNTNSKVPFFIFILA 60

 Query: 61 GMPDRILGMLITWMTKTDGQGGKRNHRFMS/TMF--IMQOKDSEVVSFGSGKDPFETRNAIL 117
 G+ DRIL L+W KTTDGQGG N+P MV+ + K+S+ + F SG+DPEE R IL
 Sbjct: 61 GVLDRLSALVWFKTTDQGGKINKPLAVTRESLTKTKAKESKEMIFDSGSDPFETRYKIL 119

```

- 60 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1411> which encodes the amino acid sequence <SEQ ID 1412>. Analysis of this protein sequence reveals the following:

-542-

Possible site: 41  
 >>> Seems to have no N-terminal signal sequence

5       ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.2905 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

10       Identities = 60/123 (48%), Positives = 82/123 (65%), Gaps = 2/123 (1%)  
       Query: 1   MLNDEDALVCDLASTYHIYDYKQLPPLKVAVPSLGLREESRINKRVISGNRVSFERRILA 60  
               M+ D+DAL CDLASTY IYDY+QLP +VAVF++GLR SRI +SG + + +LA  
       Sbjct: 1   MIKDDDALTCCLASTYGIYDYGQVAVFVAVGLRENSRIKMLSGETRALDTVLLA 60  
       Query: 61   GHFDALGHLIWMKTTDQKGNRPFEMV--STMPDHOQKDSVVSFSGCKDPFBETRNILG 118  
               G++D +L W KT DSG G+N+P+ V +       QK ++V+SF SG+DFE R +LG  
       Sbjct: 61   GIYDNTLLFWSKTKDQSGQGNKPFVSVEAISGSKSQKANDVISFVSGEDFENARKQLLG 120  
       Query: 119   FGG 121  
                   G  
       Sbjct: 121   GGG 123

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 25   vaccines or diagnostics.

#### Example 441

A DNA sequence (GBSx0478) was identified in *S. agalactiae* <SEQ ID 1413> which encodes the amino acid sequence <SEQ ID 1414>. Analysis of this protein sequence reveals the following:

30       Possible site: 36  
       >>> Seems to have no N-terminal signal sequence  
       ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.2280 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
       35       bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:ANG18636 GB:AY007505 unknown [Streptococcus mitis]  
       Identities = 40/80 (50%), Positives = 62/80 (77%), Gaps = 1/80 (1%)  
       40       Query: 3   TSSGFYKTEESRLANYELVEALADESNFLSLPKVLRLLLGQVESLQHLPLASDGTVS 62  
               TS+GF ++I + RL+NYEL+EA+++++NP LPKV++L+LG++ E LKH+R +DG V  
       Sbjct: 24   TSTGFPFETTKERLENYELLEIAISEVDTNEAVLPKVKMLGNKSEDLQGHVRTADGIVP 83  
       45       Query: 63   TEALMHEVKTIFES-QQLKK 81  
               + + E+ EIP S QLKK  
       Sbjct: 84   LDKMGAEISEIFSSQQLKK 103

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1415> which encodes the amino acid  
 50   sequence <SEQ ID 1416>. Analysis of this protein sequence reveals the following:

      Possible site: 56  
       >>> Seems to have no N-terminal signal sequence  
       ----- Final Results -----  
       55       bacterial cytoplasm --- Certainty=0.4365 (Affirmative) < succ>  
               bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
               bacterial outside --- Certainty=0.0000 (Not Clear) < succ>



-543-

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 42/75 (56%), Positives = 60/75 (80%)

Query: 2 KTTSGGFYEYKIESRLKNYELVETALDLESNPLSLPKVLRILLGDQVSESQMLRPASDXTV 61
 5 KT+SGFEY+I +RLKN+ELVET+A+ E++P ++ K++ LLLGD ++SLK H+R ++G V
Sbjct: 7 KTTSGGFETEYKPKRRLKNFLVETALKESTDPYAVYKVINLLGLDAKSLKSHVDRDABGV 66

Query: 62 STEALAESVKEIPES 76
 EA+ E+KEIPES
Sbjct: 67 DVRAIGVKEIKEIPES 81

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 442

15 A DNA sequence (GBSx0479) was identified in *S.agalactiae* <SEQ ID 1417> which encodes the amino acid sequence <SEQ ID 1418>. This protein is predicted to be Structural protein. Analysis of this protein sequence reveals the following:

```

Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3461 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:RAG18635 GB:AY007505 unknown [Streptococcus mitis]
Identities = 114/183 (62%), Positives = 142/183 (77%)

Query: 2 VANSNVTIAKFKIGGAIVTAFLTELPKDTASELANSAFKSLGYISDEGLSNEDKRESE 61
 +A +NVTIAKFKIGGA+Y+AFLGT LF D ++L++AF++LGVIS+DG++N + ESE
Sbjct: 1 MATEANVTIAKFKIGGAVYSAFLGTALPTDATTKLDQAFRALGYISDDGMNNSNPSEN 60

Query: 62 IQAWGSDVVESQAQSKADKFTYTLIEALNIEVLKEIYKGNVIGDLKGTITVYKSNKPLE 121
 I+AWGS VV S QK K D F Y LIEALN+ VLKE+YG ENV+GDL +GIT+K+NSK L
Sbjct: 61 IKAWGCVVVSQKKEKTTDFKXMLEALNLIIVLKEVYGDNVSGDLSSGITIKANSKELP 120

Query: 122 EHRLVIEMLWNNVTVKRIVIPKGVKSVGEIKYVDNBAAGYETTLQAPPAEAGNTHYEYI 181
 HCLVIE +LK +KRIVIP GKV+ + EI Y D GY TT+ AFP+A +THYEYI
Sbjct: 121 EHRLVIEITVLKGVKLRIVIPSGKVTALDEITVNDGSVLGYGTTVTAFPNAADTHYEYI 180

Query: 182 KGA 184
 KGA
Sbjct: 181 KGA 183

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1419> which encodes the amino acid sequence <SEQ ID 1420>. Analysis of this protein sequence reveals the following:

```

Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2379 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 119/182 (65%), Positives = 142/182 (77%)

```

-544-

Query: 4 NSSNVTTAKPKIGGAIYTAFLGTELPKDTASELNBAFKSLCYISEDGLSNEOKREGEIQ 63  
 ++ NVT+AKPK GGAITY+APLGTLEPKD SRLN FK+LST+SEDG+ NED R SE I+  
 5 Sbjct: 6 DTGNTVTEAKPKTGGAIYAPLGTLELPKDAKSELNITFKMLGYVSGDGVNEDTRSENIEK 65

Query: 64 AWGGD+VBSAQSKADKFTYTLIEALNIHVLEKIYGDVITGDLKTCGIVKENS/KPLEEH 123  
 AWGGD+V + Q K DRFTY LIE+LH+HVLKE+YG NVIGDL GI +KSNK LE H  
 10 Sbjct: 66 AWGGDIVGAVCTEKSDKFTYKLTIESLNVLEKIVGAVNVITGDLGGGIQIKSNS/LELAH 125

Query: 124 CLVIRMLNNTVVKRIVIPKGVSEVGIKIVDNEAGYETTLQAFDAGNTHYVYIKG 183  
 +V++MI+ +KRIV+P KV EVGRKIYVD E GYETTL+ FFD +G+TH EYI  
 15 Sbjct: 126 VIVVDMINAGSGLKRIVLENAKVDVSGIKIYVDGEVVGYYETTLKCFPDKDGUTHRYIVK 185

Query: 184 AG 185  
 G  
 15 Sbjct: 186 FG 187

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 20 Example 443

A DNA sequence (GBSx0480) was identified in *S. agalactiae* <SEQ ID 1421> which encodes the amino acid sequence <SEQ ID 1422>. Analysis of this protein sequence reveals the following:

Possible site: 58  
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2214 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

30 The protein has homology with the following sequences in the GENPEPT database:  
 >GP:AA818710 GB:U38906 ORF35 [Bacteriophage rlt]  
 Identities = 52/78 (66%), Positives = 66/78 (83%)

35 Query: 1 MSKPKFKLNKAGVAELMKSSSEMQVLTTKATAIRERCGDGYAQDIHVGNRANAMVSAKT 60  
 M+K FKLN++GVA +MKS EMQ +L KA++RCS GY QD+HVGKRNANAMV A+T  
 Sbjct: 1 MAENLPKLNRSQVASHMMKSPMQAILKEKASAVKRCGPGYQGMHVGNRANAMVFAET 60

Query: 61 IKAKKKNKQNTILKAVR 78  
 +AK+DN KQNT+LKAVR  
 40 Sbjct: 61 YQAKRDNMQNTILKAVR 78

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1423> which encodes the amino acid sequence <SEQ ID 1424>. Analysis of this protein sequence reveals the following:

45 Possible site: 54  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2446 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

50

An alignment of the GAS and GBS proteins is shown below:

Identities = 75/78 (96%), Positives = 76/78 (97%)

55 Query: 1 MSKPKFKLNKAGVAELMKSSSEMQVLTTKATAIRERCGDGYAQDIHVGNRANAMVSAKT 60  
 MSKPKFKLN+AGVAELMKSSSEMQVLTTKATAIRERCGDGY QDIHVGNRANAMV KT  
 Sbjct: 1 MSKPKFKLNKAGVAELMKSSSEMQVLTTKATAIRERCGDGYQDIHVGNRANAMVSTKT 60

-545-

Query: 61 IKAKKDNKSNITLLGAVR 78

IKAKKDNKSNITLLGAVR

Sbjct: 61 IKAKKDNKSNITLLGAVR 78

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 444**

A DNA sequence (GBSx0481) was identified in *S.galactiae* <SEQ ID 1425> which encodes the amino acid sequence <SEQ ID 1426>. Analysis of this protein sequence reveals the following:

10 Possible site: 38  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.2888(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

20 >GP:AB18709 GB:U38906 ORF34 [Bacteriophage xlt]  
Identities = 41/59 (69%), Positives = 45/59 (75%)

Query: 1 MTGKKVEYILAIPKGDHNDKEVCFDDKKWTVGLALBGIIEGLIPLENNKKVMVERY 59  
+TGKK Y LAIPK D HDNE+K+V FF K WET G LGIEI LIPL+NNKKV VE Y  
25 Sbjct: 56 LTGKKALYTLAIPKGDTHDNNKKVRFFGKWTVPGLBGIIEGLIPLENNKKVTVERY 114

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1427> which encodes the amino acid sequence <SEQ ID 1428>. Analysis of this protein sequence reveals the following:

30 Possible site: 39  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.2779(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 51/60 (85%), Positives = 57/60 (95%)

40 Query: 1 MTGKKVEYILAIPKGDHNDKEVCFDDKKWTVGLALBGIIEGLIPLENNKKVMVERY 60  
+TGKKVEY+LAIPKGD+HDNE+KEV FF KGWTVG+ LGIEIIEGLIPL+NNKKVMVERY  
Sbjct: 50 LTGKKVEYVLAIPKGDHNDNNKKVRFFGKWTVPGLBGIIEGLIPLENNKKVMVERY 109

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 445**

A DNA sequence (GBSx0482) was identified in *S.galactiae* <SEQ ID 1429> which encodes the amino acid sequence <SEQ ID 1430>. Analysis of this protein sequence reveals the following:

50 Possible site: 25  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2770(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

```

5 >GP:AB18708 GB:U38906 ORF33 [Bacteriophage r1c]
 Identities = 89/130 (68%), Positives = 106/130 (81%), Gaps = 1/130 (0%)

 Query: 1 MNFATTDVILLWRQLSVDEIKRABALLETVSDTLRLKASKVGNLDEMILETP-YPAT 59
 M FAT DD+ +LWR L DE +RAS LLS VSD+LR EA KVG++L MI E P YPA+
 Sbjct: 1 MNFATTDVILLWRQLSVDEIKRABALLETVSDTLRLKASKVGNLDEMILETP-YPAT 60

10 Query: 60 VLKSVTVDIVARTLMTATQGBPMQSQSALGYTWSGTYLVPQGLFIKDSKRLGLK 119
 V+KSVTVDIVARTLMT+T EPM+Q ++SALGY+ SG+YLVPQGLFIK+SEL RLGLK
 Sbjct: 61 VVKSVTVDIVARTLMTSTQCEPNTQTTSEALGYSVSGSYLVPGGLFIKNSLRLGLK 120

15 Query: 120 QRYGGIELYG 129
 QR+G I+ YG
 Sbjct: 121 QRYGVIDFYG 130

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1431> which encodes the amino acid sequence <SEQ ID 1432>. Analysis of this protein sequence reveals the following:

```

20 Possible site: 37
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2061(Affirmative) < succ>
25 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

30 Identities = 116/138 (84%), Positives = 129/138 (93%)

 Query: 3 NFATTDVILLWRQLSVDEIKRABALLETVSDTLRLKASKVGNLDEMILETPYPATVLK 62
 NFATTDVILLWR LSVDE+KRA ALL+ VSDTLR+EA KVGK+LD+ +++ PTF V+K
 Sbjct: 3 NFATTDVILLWRQLSVDELKRRANALLKVVSDTLRMEADKVGEDLQKTMVDKPYFVNVIK 62

35 Query: 63 SVTVDIVARTLMTATQGBPMQSQSALGYTWSGTYLVPQGLFIKDSKRLGLKQRY 122
 SVTVDIVARTLMT+T+GSPM+QSCSALGYTWSGTYLVPQGLFIKDSKRLGLKQRY
 Sbjct: 63 SVTVDIVARTLMTSTRGEPMAQSQSALGYTWSGTYLVPQGLFIKDSKRLGLKQRY 122

40 Query: 123 GGIELYGEIERNSYFSR 140
 GGIELYGEIER-NS FSR
 Sbjct: 123 GGIELYGEIERNSCFPSR 140

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### 45 Example 446

A DNA sequence (GBSx0483) was identified in *S.galactiae* <SEQ ID 1433> which encodes the amino acid sequence <SEQ ID 1434>. This protein is predicted to be Structural protein. Analysis of this protein sequence reveals the following:

```

50 Possible site: 30
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3015(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
55 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

-547-

>GP:ABE18706 GB:U38906 Structural protein [Bacteriophage rlt]  
 Identities = 132/296 (44%), Positives = 189/296 (63%), Gaps = 8/296 (2%)

Query: 5 IKAOTLFKPELVTEIMSKVKHSTLAKLGGQTFIPFNGVQFVFNLDGNAQIVGSEBQL 64  
 + OTLF P LVT+++SKV G S++AA+LS Q PIPFNG + F F +D +V E +K  
 5 Sbjct: 3 LNKGLTFDPTLVTDLISKVAGESIALRKAQKPIFNGEKVPTFMDSEIDVVAESGKKT 62

Query: 65 GNTAKVTSKIIKPLKPVYQARMTDEFKYASEEKRILFKHYADGFPAKMAEAFDIAIHG 124  
 + + + P+K Y AR++DEF YAS+E++N L+ + DGFPAK+A D+ A HG  
 10 Sbjct: 63 HGGVTLAQQTMVPIKVEYGARISDEFPMIAGDEEIKNLLQSFNDGFAKIKVARGIDLMAFHG 122

Query: 125 LBPETMTDTSFPAKATNSFGVVTGNNVIKYAEK--IDN--IDAAVTTIVANGNDVGTIAL 180  
 + EP T ++ TN FD VT K EA + D N I+ AV + DVTGIA+  
 15 Sbjct: 123 VNEPLGTASAVIGTNHFDKSVTQ---KVEAPRGTDAPNGAENAVELLTGVDADVGTIAI 179

Query: 181 SPQAGQMSKRRDKKFDNVMYPERFQGRPSNFFNMILDINKTLTMKGQTAKDDBAIVG 240  
 +P ++K+KD EN ++PE +G P + +D+NKT++ T + D AI+GDF  
 20 Sbjct: 180 HPSFRSALAKQKDLQNALFPELKGATPOTINGLPVDVNTKVSOMSLQR-DRAIIGDF 238

Query: 241 QMFNFKGAYARNIPMEIIIEYGDPDGSGRDLKAYNEILLRTEAFIGWGILDKAFSRV 296  
 N FNGYA +P+E+I+YGDPD SG DLK YN++ R E P+GWGILD F+RV  
 25 Sbjct: 239 ANGFWGAYAEVFEVLEVICYGDPTNSGLDLKYNQYIYRAELFLGWGILDATFARV 294

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1435> which encodes the amino acid sequence <SEQ ID 1436>. Analysis of this protein sequence reveals the following:

Possible site: 30  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.2772 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 133/298 (44%), Positives = 187/298 (62%), Gaps = 2/298 (0%)

Query: 1 MAESIKAOTLFKPELVTEIMSKVKHSTLAKLGGQTFIPFNGVQFVFNLDGNAQIVGSEB 60  
 M +LF LV+++++KVKHSA+LAKLS Q PIPFNG ++F F LD + +V E  
 35 Sbjct: 1 MGTETSKASLFDKHLVSLINKVKHSSLAKLSQKPIPFNGSKKEPTFTLSDIDVVAEN 60

Query: 61 EQKLGNTAKVTSKIIKPLKPVYQARMTDEFKYASEEKRILFKHYADGFPAKMAEAFDIA 120  
 +K + I P+K Y AR++DEF YA+ES+++ LK + +GFAK+A D+  
 40 Sbjct: 61 GKKTGGLLEPVTVPPIKVEYGARLSDEFLYATBEERKIDILKAFNGFAKLARGIDIL 120

Query: 121 AIHGLEPRTMTDTSFPAKATNSFGVVTGNNVIKYAEKIDNDIDAAVTTIVANGNDVGTIAL 180  
 A+HG+ PRT + TN FD VT V E++ D NI+RAV I + VTG+A+  
 45 Sbjct: 121 AMHGINFRITKASDVIGTNHFDKSVTVVVKFTSEADANIRAAVNLQSGEVTGLAM 180

Query: 181 SPQAGQMSK-RDKKFDNVMYPERFQGRPSNFFNMILDINKTLTMKGQTAKD-DHAIVG 238  
 + ++K + MYPE +G P + + +N T+ A+ D I+G  
 50 Sbjct: 181 DTEFSTALAKVTNGEMCPKMYPELAWGANEDSINGLSVNTTVGAGDAESKDLIVIG 240

Query: 239 DFQNMFKGAYARNIPMEIIIEYGDPDGSGRDLKAYNEILLRTEAFIGWGILDKAFSRV 296  
 DF++MFKGAYA+ IPMEII+YGDPD SG+DLK YN+I LR EA+IGWGILD K+P+RV  
 55 Sbjct: 241 DFESMFKGAYARQIPMEIIKYGDPTNSGKDLKYNQYIYRAEYIGWGILDKAFSRV 298

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 447

60 A DNA sequence (GBSx0484) was identified in *S.agalactiae* <SEQ ID 1437> which encodes the amino acid sequence <SEQ ID 1438>. Analysis of this protein sequence reveals the following:

-548-

Possible site: 61  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 5           bacterial cytoplasm --- Certainty=0.2224 (Affirmative) < succ>  
            bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
            bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9659> which encodes amino acid sequence <SEQ ID 9660> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AB18705 GB:U38906 CRF30 [Bacteriophage r1t]  
Identities = 64/158 (40%), Positives = 101/158 (63%), Gaps = 8/158 (5%)

- 15   Query: 43 MSEFKVIETQBELDTIVKARIARERE----KYQYDQDKTRVEKLETNSSSLQTLNDAK 98  
          MSE + +TQBEL+ I++ R+AR++E   + DVD+LKT++ LE +H++ Q + ++K  
      Sbjct: 1 MSENILPKTQBELNQIITRLARQKETIEANFADYDEKTKIAALEADNTAYQATIEBSK 60
- 20   Query: 99 SNTDSYTEKITTLENQIAGYEANLRTKVALQYGLPIDLANRLQGDDDEGLKVDASRLAS 158  
          S   + ++   E QI+GY+ I+ +A++ GLP+DLA+RL GDDE+ LK DAER +  
      Sbjct: 61 S----WBOEKADYKQKQISGYKTTLQKQSIATKAGLPLDLADRLSGDDEESLKADAEKPSG 116
- Query: 159 FIKPQSQPPPTKSENEPIITDQKRWGIMARNLVNKEG 196  
          FIKP P P K EP + D K+ + ++ L +GE  
25   Sbjct: 117 FIKPKTTPAPLKDVERNLGDGKGAYRKLVDGLKTEGE 154

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1439> which encodes the amino acid sequence <SEQ ID 1440>. Analysis of this protein sequence reveals the following:

Possible site: 59  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 30           bacterial cytoplasm --- Certainty=0.3476 (Affirmative) < succ>  
            bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
35           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 128/149 (85%), Positives = 136/149 (90%)

- 40   Query: 43 MSEFKVIETQBELDTIVKARIAREKQYQYDQDKTRVEELETENSSSLQTLNDAKSNITD 102  
          MSEFKVIETQBELDTIVKARIAREKQYQYDQDKTRVEELETENSSSLQTLNDAKSNITD  
      Sbjct: 1 MBSFKVIETQBELDTIVKARIAREKQYQYDQDKTRVEELETENSSSLQTLNDAKSNITD 60
- 45   Query: 103 SYTEKITTLENQIAGYEANLRTKVALQYGLPIDLANRLQGDDDEGLKVDASRLASFIKP 162  
          SYTE+I+TL+NQIA YE ANLRTKVALQYGLPIDLA+RLQGDDDEGLKVDASRLASFIKP  
      Sbjct: 61 SYTESISTLRKQIADYETANLRTKVALQYGLPIDLADRLQGDDDEGLKVDASRLASFIKP 120
- Query: 163 SQQPFPPTKSENEPIITDQKRWGIMARNL 191  
          SQQPFP KSENEP I   +A + + + L  
50   Sbjct: 121 SQQPPEAKSENEPNIDSNADANYRALVQGL 149

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 448

- 55 A DNA sequence (GBSx0485) was identified in *S.agalactiae* <SEQ ID 1441> which encodes the amino acid sequence <SEQ ID 1442>. Analysis of this protein sequence reveals the following:

Possible site: 56

-549-

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

- 5           bacterial cytoplasm --- Certainty=0.2888 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

- 10 >GP:AAB18704 GB:U38906 ORF29 [Bacteriophage rlt]  
 Identities = 322/461 (69%), Positives = 383/461 (82%)
- 15 Query: 8 KLGQRPTQSVNLHFAKTLAHEAINYYKKTGLSCYLWQENMLIPMAINKRNLNVHVKYQ 67  
           + GNQ PTQSV L P +T RAI Y+K+ CY WQ+N+L +MAI+ED LN HQR+G  
       Sbjct: 6 RFRQYPTQSVILPFPFTKTYQRATEIYKSKHSCYPWQKNLLKEVMAIDEDGLNTHQK 65
- 20 Query: 58 YAIPIRRNGKTEVVIYILMALHKGKLTILTAHRISTSHSSPEKVKYKYLMSGYVDGEDFI 127  
           Y+IPRRNGKTE+VYILSL+L +GL ILTAHRISTSHSS+EK+KYLE SGYV+GEDF  
       Sbjct: 66 YSIPRRNGKTEIVYILMSLWQLSLILTAHRISTSHSSYELKKYLEDGQYVEGEDFK 125
- 25 Query: 128 SNKAGQQRERIEFKSGSVIQFRTRTSNGLGEGFDLLIDEACYEYTBQESALKYTVTDS 187  
           S KAGQQR+E SG VIQFRTRTS+GGLGEGFD+L+IDEACYEYTBQESALKYTVTDS  
       Sbjct: 126 SIKAGQERLELIESGGVIQFRTRTSNGLGEGFDILVIDEACYEYTBQESALKYTVTDS 185
- 30 Query: 188 DNPMTIMCGTPTPTMVSTGTVFESYRKCGLKDRRYSQWAEWSVDEMOPHDVSKWYVNP 247  
           DNPMTIMCGTPTPTVS+GTVF +YR + G +YSQWAEWSV++++ IHDV++WY +NP  
       Sbjct: 186 DNPMTIMCGTPTPTFVSSTGYFTNYRDNITAGKAYSQWAEWSVEDVIDHDVEAWYNPN 245
- 35 Query: 248 SMGYHLNERKTEAZLGEDEIDHNIQRLQYWPSPNQKSVISEKEWAKLVQVPELKSCLF 307  
           SMGYHLNERKTEAZLGEED+EDH+QRLGYWP +NQKSVISE+EW LKV ++P +K LFL  
       Sbjct: 246 SMGYHLNERKTEAZLGEDEIDHNIQRLQYWPYKYNQSVISEQEWNAKLVNRLPVEK 305
- 40 Query: 308 VGKIFGQDGENUSLSIARASENKVVFVAIDCLSVRNGTQWIINFLSADIARVVDGSL 367  
           VGK+G DG HV++SIA + KVFVE IDC S+RNG QWIINFLK AD+ KVV+DG S  
       Sbjct: 306 VGKIFGNDGANVMSIAVKTLSGKVFVETIDCSIRNGH+WIINFLKGDVKEVVDGSG 365
- 45 Query: 368 GQELLAQENRSHGLKLPKVAEITANTMWEQGMQSTICHNDQPSLTAVVNTCEKRG 427  
           GQ +L EM++ LK+P LP V EII AR++WEGGI Q+ CH+ QPSL+ VVNTC+RR  
       Sbjct: 366 GQILTSBMDPKLKEFILPTVKEINRNSLWEGGIFQNPCHSGQPSLTAVVNTCDKRG 425
- 40 Query: 428 IGSNGGFGYKSLYDQDRDISLWDSALLAHWICYTTKPKRQR 468  
           IG++GGFGYK +DD DISLWDSALLAHW C KPK+KQ+  
       Sbjct: 426 IGTSGGFGYKSGFDMDISLWDSALLAHWCSNNKPKKKQ 466

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1443> which encodes the amino acid  
 45 sequence <SEQ ID 1444>. Analysis of this protein sequence reveals the following:

Possible site: 32

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

- 50           bacterial cytoplasm --- Certainty=0.3133 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

- 55 Identities = 437/471 (92%), Positives = 459/471 (96%)
- 60 Query: 1 MVTIKTKLGNQRPTQSVNLHFAKTLAHEAINYYKKTGLSCYLWQENMLIPMAINKRNL 60  
           MVTIKTK KLGQRPTQSVNLHFAK+LAHEAINYYKKTGLSCY WQ NMLIP+MAI+P+ L  
       Sbjct: 6 MVTIKTKLGNQRPTQSVNLHFAKSLAHEAINYYKKTGLSCYFWQENMLIPMAIDEDKGL 65
- 60 Query: 61 WYHVKQGYAIPIRRNGKTEVVIYILMALHKGKLTILTAHRISTSHSSPEKVKYKYLMSGY 120  
           WYHVKQGYAIPIRRNGKTEVVIY+LMALHKGKLTILTAHRISTSH+SPKVKYKYLMSGY  
       Sbjct: 66 WYHVKQGYAIPIRRNGKTEVVIYVQLMALHKGKLTILTAHRISTSHSSPEKVKYKYLMSGY 125

-550-

Query: 121 VDGRDFISNKAQGQRIEFKSGSGSVIQFRTIRTSNGGLGSGFDLLIIDBAQETTABQESAL 180  
 VDGRDFISNKAQGQRIEFK+SG+VIQFRTIRTSNGGLGSGFDLLIIDBAQETTABQESAL  
 Sbjct: 126 VDGRDFISNKAQGQRIEFKAGGAVIQFRTIRTSNGGLGSGFDLLIIDBAQETTABQESAL 185

5 Query: 181 KYTVTDSDEHFTIMOGTPPTMVSTGTVFVSYSKRLKGDORRYSQWABMSVDEMOPHIDVK 240  
 KYTVTDSDEHFTIMOGTPPTMVSTGTVFVE+YRK+CLKG++RYSQWABMSV EM I+DV  
 Sbjct: 186 KYTVTDSDEHFTIMOGTPPTMVSTGTVFVAYRKDKCLKGNGRYSQWABMSVPEHMKINDVS 245

10 Query: 241 SWYVNPSSNGYHLEKRIKRLGDRIDHNIQRIQYWPSPNQKSVISRKSWAKLKVEQVP 300  
 SWY++NFSMG+HLNERKIRLDELGDIDHNIQRLQYWPSPNQKSVISRKSWAKLKVEQVP  
 Sbjct: 246 SWYISNFSMGPHLNERKIRLDELGDIDHNIQRLQYWPSPNQKSVISRKSWAKLKVEQVP 305

Query: 301 ELKSKLFVGIKFGQDGNVLSIAAPASENKVFVEAIDCLSVRNGTQWTLINFLKSADIAT 360  
 ELKSKLFVGIKFGQDGNVLSIAAR SENKVFVE IDCLSVRNGTQWTLINFLKSADIAT  
 Sbjct: 306 ELKSKLFVGIKFGQDGNVLSIAARTSENKVFVETIDCLSVRNGTQWTLINFLKSADIAT 365

15 Query: 361 VVVDGASQGLLAQEMRHHGLKPELPKVAEIIITANMWSQIMQETICHNDQPSLTAVV 420  
 VV+DGASQGLLAQEM++ GLKKPELPKVAEIIITANMWSQIMQETICH+DQPSLTAVV  
 Sbjct: 366 VVIDGASQGLLAQEMKDQGLKKPELPKVAEIIITANMWSQIMQETICHSDQPSLTAVV 425

20 Query: 421 TNCKRKQISNGGFGYKSLYDORDISLMSALLAHWICTTTKPKRKQRTSC 471  
 TNCKRKQISNGGFGYKSLYDORDISLMSALLAHWICTTTKPKRKQRTSC  
 Sbjct: 426 TNCKRKQISNGGFGYKSLYDORDISLMSALLAHWICTTTKPKRKQRTSC 476

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 249

A DNA sequence (GBSx0486) was identified in *S.galactiae* <SEQ ID 1445> which encodes the amino acid sequence <SEQ ID 1446>. Analysis of this protein sequence reveals the following:

30 Possible site: 32  
 >>> Seems to have no N-terminal signal sequence

---- Final Results ----

35 bacterial cytoplasm --- Certainty=0.2745 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 250

A DNA sequence (GBSx0487) was identified in *S.galactiae* <SEQ ID 1447> which encodes the amino acid sequence <SEQ ID 1448>. Analysis of this protein sequence reveals the following:

45 Possible site: 32  
 >>> Seems to have no N-terminal signal sequence

---- Final Results ----

50 bacterial cytoplasm --- Certainty=0.2568 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AB18703 GB:U38906 ORF28 [Bacteriophage x1t]



-551-

Identities = 124/250 (49%), Positives = 164/250 (65%), Gaps = 3/250 (1%)

Query: 2 VDDVLPKLLKSVQDFFKHHFGSEVVAKAFAPLQAKKATYKTVNEFAVEVGRLLSLALAN 61  
 ++D+LP LL++ QDF++ S+ + ++ L+ KRAIT NEF VEVG++LS L  
 5 Sbjct: 1 MEDILPFLLEKINQDFBRRAANSKLLQSMELLCTKKATYIQANFPGVVEVQGLSDVIAT 60

Query: 62 SVISDELDPGKMYNIANRLVNDTLRHNYKLISDYAGVQCNLAKQAKISLKIQRPPINQ 121  
 V D LPDGRMY+NIA+RI+N L+ H+ LIS Y+ DVQ LN+ A LK Q P IANQ  
 10 Sbjct: 61 HVTVDVLPDGRMYNIALADRLASLILKKNFDLISGYSDVQSELANQLAGFKLKSQVPRINQ 120

Query: 122 DKIDGLVNRILASEPVFDVVKLLDEPVIWFSQISVDDCIRANADPHFKTGLKPTIERIST 181  
 D+IDG+VNR++SS F+ + WLL EPIV FSGS+VDD ++ N D<sup>E</sup> K GLKP I R  
 Sbjct: 121 DRIDGIVNRISSEDDFEKILMLLEKPIVTFVSQSVDDTLKKNIDPQAKAGLKPPIVRKLIV 180

Query: 182 GKCCDWCDRLAGRYVHEEPKDFYKHHQHCQCVIDYHPK--NKKRQNSWSKKWKEITYDI 239  
 GK CDWC LAG Y Y E D Y RH+ C+C ++Y P+ + KRQ+ WSK W D  
 15 Sbjct: 181 GKACDWCRNLGSDYTFVPSDVYRHERCRCTVSDYDPDIDKKGQDWSKQWDPDKDA 240

Query: 240 -LERRKQMT 248  
 + RK +H+  
 20 Sbjct: 241 KIAERKRLML 250

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1449> which encodes the amino acid sequence <SEQ ID 1450>. Analysis of this protein sequence reveals the following:

Possible site: 32  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3099 (Affirmative) < succ>  
 30 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 169/261 (64%), Positives = 207/261 (78%), Gaps = 2/261 (0%)

Query: 1 MYDDVLPKLLKSVQDFFKHHFGSEVVAKAFAPLQAKKATYKTVNEFAVEVGRLLSLALAN 60  
 MYDDVLPKLLKSV+QDFEK+FG+S+VV KAPALQAKK TYKTVNEFA+EVGRLLSLAL  
 35 Sbjct: 1 MYDDVLPKLLKSVQDFFKYPGSEVSVVTKAPALQAKKTYKTVNEFAIEVGRLLSLALT 60

Query: 61 NSVISDELDPGKMYNIANRLVNDTLRHNYKLISDYAGVQCNLAKQAKISLKIQRPPINQ 120  
 SV SD+LPDGRMYNIA RL++T+ NYKLIS YAGDQ+ LN+ A+I LK+QRPPIN  
 40 Sbjct: 61 GSVSSDKLPDGRMYNIAKRLDDETWGRNYKLISGYAGVQRLINENQIGLKVORPPIN 120

Query: 121 DKIDGLVNRILASEPVFDVVKLLDEPVIWFSQISVDDCIRANADPHFKTGLKPTIERIST 180  
 +DKI+G+VNL SE FDDVKNL EPIVWFSQISVDD I+ANAD +KTG+ P + R  
 45 Sbjct: 121 RDKINGVNRILDEENTYDDVKNLFGREPIWFSQISVDDTIKANADLQYKYGMTPOVVRTE 180

Query: 181 TKGCCDWCDRLAGRYVHEEPKDFYKHHQHCQCVIDYHPKNGKQNSWSKKWKEITYDI 238  
 +G CC+WC + G Y Y + PKD ++RHQ C+C +DY PKNGK Q++WSK W K +T+  
 50 Sbjct: 181 SGNCCEWCREVGTYSYPKVPKVDVNRHRQRCRCTLVDYPKNGKVGQSMKIVRKKKTKYE 240

Query: 239 ILERRKQMDIRDNRRKSDI 259  
 +ER ++ + K+DI  
 55 Sbjct: 241 SIERVEKFKESALVSSIKNDI 261

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-552-

## Example 451

A DNA sequence (GBS0488) was identified in *S.galactiae* <SEQ ID 1451> which encodes the amino acid sequence <SEQ ID 1452>. This protein is predicted to be Structural protein. Analysis of this protein sequence reveals the following:

```

5 Possible site: 58
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.38 Transmembrane 93 - 109 (93 - 110)

10 ----- Final Results -----
 bacterial membrane --- Certainty=0.1553 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15 >GP:ZAC39307 GB:AF022773 ORF5 [Lactococcus bacteriophage phi31]
 Identities = 271/410 (66%), Positives = 326/410 (79%), Gaps = 2/410 (0%)

Query: 1 MNTMGMGLQRLKALFKTGVDKRYRYAMDDRDNRISIVMPDNVREMYRSVBIWTKAGVD 60
 M G+G+L+ K++ K + RY YAM D + I +P + + YRS++ W AKGVD
20 Sbjct: 1 MTEKGIGYLRFKLSVHKRRARMRYEQYAMKQVDRFKGIIPTQALSCQYRSILGWCAKAGVD 60

Query: 61 SLADRIIFREFANDDPNAWEIKANINPDIFPDATQSAIASCFCFYIMPCKEDSLPMKQ 120
 SLADR+IFREF NDDF EIF+ NNDIFSD+A+ SALIASC F+YI G+ D++ ++Q
25 Sbjct: 61 SLADRLIFREFENDDFTVNRIPEKNINPDIFDQAVLSALIASCSFYISKGENDAV-RLQ 119

Query: 121 VIEASKATGILDPTFTLLTBGYAVLESDSNENPTLEAYPTGEKTYWPKDEK-YSIDNS 179
 VIEA ATGI+DF T LTBGYAVLE D N N LEA+F ++T YY +D + SI N
30 Sbjct: 120 VIRAVNATGIDITPTGLTBGYAVLERDENNVVLEAHFLPRTDYYYRDSRNIISANF 179

Query: 180 TGHPELLVPVIHRPDVAVRPFGRSRITKAGMYHQAAKKTLERAETVIAFYSPFQKYVLMD 239
 TGHPELLVP+IHRPDVAVRPFGRSRIT++GM Y Q AKRTLERA+VIAEFYSPFQKY G+
30 Sbjct: 180 TGHPELLVPVIHRPDVAVRPFGRSRITRSGMYQSNAAKTLERAQVIAFYSPFQKYVTGLS 239

Query: 240 PDAEPMKQRAVTVSTLLEISKDEGDKPTVGQFTASMAFFMDLKMAYSLFAGGSGILT 299
 DAEPM E W+ATVS++L+ +KDEGDKPT+GQPT SM+FF + L+ A+ FAG +GLTL
35 Sbjct: 240 DDAEPMETNKATVSSMLQFTKDEGDKPTLQQTQPSMSPFTEQLRTAAAGFAGETGLTL 299

Query: 300 DDIGFPSPDNPSSVEAIKAHENLRAAGKKAQRFSFSSGFIAVAYIAVCLRDDFFYLENQM 359
 DDIGFP SINTPSSVEAIKA+HENLR AGRRAQRS +G LNAV+A CLRDD PYLR QF
40 Sbjct: 300 DDIGFPVSDNPSSVEAIKASHENLRLAGKKAQRSLGAGLIANVAYLAACLRDDVPTLREQFS 359

Query: 360 DTEIKWEPLFEADANMLTLVGDSAIKLQAIPGFMDADVIRDLTGKVGSD 409
 T+ KWEPLFEADA+ML+L+GDGAIKLQAIP F++ D IRDLTG+KG++
45 Sbjct: 360 KTKPKWEPLFEADASMLSLIGDGAIKLQAIPFINKDTIRDLTGIGKAE 409

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1453> which encodes the amino acid sequence <SEQ ID 1454>. Analysis of this protein sequence reveals the following:

```

50 Possible site: 58
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.38 Transmembrane 93 - 109 (93 - 110)

55 ----- Final Results -----
 bacterial membrane --- Certainty=0.1553 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

 Identities = 395/422 (93%), Positives = 407/422 (95%)

60 Query: 1 MNTMGMGLQRLKALFKTGVDKRYRYAMDDRDNRISIVMPDNVREMYRSVBIWTKAGVD 60
 MNTMGMGL+KRLALFKTGVDKRYRYAMDDRD+TRISIVMP+NVREMYRSV+BIWTKAGVD

```

-553-

Sbjct: 1 MNYNGMYLRRKLALFKTGVDKRIYYAMDRODTRSIIMPNNVRNMYRSLVSWTAKGVD 60

Query: 61 SLADRIIFREFANDDFNWEIFKANNFDIFFDTAIQSALIASCCFYIIMPGEISLPSMQ 120  
SLADRIIFREF NDDFNWEIFKANNFDIFFDTAIQSALIASCCFYIIMPGE ED LPSMQ

5 Sbjct: 61 SLADRIIFREFPTNDDFNWEIFKANNFDIFFDTAIQSALIASCCFYIIMPGEISLPSMQ 120

Query: 121 VIEASKATGILDPFTFLITGCVAVLESDESNINPTLRAYPTGRKTYWYFDKPKYSLINST 180  
VIEASKATGILDPFTFLITGCVAVLESDESNINPTLRAYPTGRKTYWYFDKPKYSLINST

10 Sbjct: 121 VIEASKATGILDPFTFLITGCVAVLESDESNINPTLRAYPTGRKTYWYFDKPKYSLINST 180

Query: 181 GHPLLPVPIHRDAVRPFGRSRIITKAGMYHQKAAKTLERAEVTARFYSFPQKYVLGMDP 240  
GHPLLPVPIHRDAVRPFGRSRIITKAGMYHQKAAKTLERAEVTARFYSFPQKYVLGMDP

Sbjct: 181 GHPLLPVPIHRDAVRPFGRSRIITKAGMYHQKAAKTLERAEVTARFYSFPQKYVLGMDP 240

15 Query: 241 DARPMENRATVSTLLRISKDEDDGDKPTVGQFTTASMAFFMHLQYASLFAQGSGLTLD 300  
DARPMENRATVSTLLRISKDEDDGDKPTVGQFTTASMAFFMHLQYASLFAQGSGLTLD

Sbjct: 241 DARPMENRATVSTLLRISKDEDDGDKPTVGQFTTASMAFFMHLQYASLFAQGSGLTLD 300

20 Query: 301 DLGFPDSNPFSSVEIKAAHENLRAGRKACRSFSSGFLNAVYAVCLRDFFYLRLQFMD 360  
DLGFPDSNPFSSVEIKAAHENLRAGRKACRSFSSGFLNAVYAVCLRDFFYLRLQFMD

Sbjct: 301 DLGFPDSNPFSSVEIKAAHENLRAGRKACRSFSSGFLNAVYAVCLRDFFYLRLQFMD 360

Query: 361 TEIKWEPLFEADANMLTAVGDAIKLQAIQGFMDADVIRDLTGVGKSDNPIPKATEVTT 420  
TEIKWEPLFEADANMLTAVGDAIKLQAIQGFMDADVIRDLTGVGKSDNPIPKATEVTT

25 Sbjct: 361 TEIKWEPLFEADANMLTAVGDAIKLQAIQGFMDADVIRDLTGVGKSDNPIPKATEVTT 420

Query: 421 DG 422  
DG

30 Sbjct: 421 DG 422

SEQ ID 1452 (GBS364) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 73 (lane 6; MW 50kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 11; MW 75kDa).

GBS364-GST was purified as shown in Figure 216, lane 10.

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 452

A DNA sequence (GBSx0489) was identified in *S. agalactiae* <SEQ ID 1455> which encodes the amino acid sequence <SEQ ID 1456>. Analysis of this protein sequence reveals the following:

40 Possible site: 16  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.4063 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

50 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1457> which encodes the amino acid sequence <SEQ ID 1458>. Analysis of this protein sequence reveals the following:

Possible site: 16  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

55 bacterial cytoplasm --- Certainty=0.4120 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

-554-

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

An alignment of the GAS and GBS proteins is shown below:

Identities = 101/118 (85%), Positives = 110/118 (92%)

5           Query: 1   MKKKCLICKKTFOAKTNRSLYCSBCKRKGIREQKQKMKQKRADKKQEKIKVLNADV 60  
               +KKKCLICKK FQAKTNR+LYCSBCKRKG REQKRLMKQKRA+++KEK KVLN N DV  
               Sbjct: 1   LKKKCLICKKNFQAKTNRTLYCSBCKRKGREKQKRLMKQKRABQRREKKKVLNPNVDV 60

10          Query: 61   TEKPKKIRGLVQHYKKLKEILLNBSBFGFTGIALVSGDIHERNFVDLVMQKIKBQQ 118  
               TEKPKKIRGL QHYK/LK+EIL NBSBFGFTGI L+BGID+HEENFVDLVMQKIKBQ+  
               Sbjct: 61   TEKPKKIRGLAQHYKKLKEILLNBSBFGFTGITLIEGIDVHERNFVDLVMQKIKBQK 118

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 15 vaccines or diagnostics.

**Example 453**

A DNA sequence (GBSx0490) was identified in *S.agalactiae* <SEQ ID 1459> which encodes the amino  
 acid sequence <SEQ ID 1460>. Analysis of this protein sequence reveals the following:

Possible site: 19

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0633(Affirmative) &lt; succ&gt;

bacterial membrane --- Certainty=0.0000(Not Clear) &lt; succ&gt;

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

The protein has homology with the following sequences in the GENPEPT database:

&gt;GP:AA039305 GB:AF022773 ORF3 [Lactococcus bacteriophage phi31]

Identities = 75/109 (68%), Positives = 87/109 (79%), Gaps = 1/109 (0%)

30          Query: 29   LRADKKGTGTHRVAFENKRRLLKTAHLGICGRFVDKSLKYPHPLSAADHIVPIAKGGH 88  
               LRAD+ G HRVAF+IK++ LKLT + CSIOG+P+DK LK P PLS +DHI+PI RGGHP  
               Sbjct: 3   LRADRTGAHRVAFDKNRKLILKLTQNTGICGKPIDKRLKAPDPLSPVVDHITPIPIAKGGH 62

35          Query: 89   SSIDNLQLTHWQCNQKSDKLFQNTAVRATVVGNNRLPQSRDWSYS 137  
               S++DNLQL HW CNRQKSDKLF N    V+GNRNLQPSRDWSYS S  
               Sbjct: 63   SAMDNLQLAHWTGNRQKSDKLF-HVKQEEPKVLGNRNLQPSRDWSYSVS 110

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1461> which encodes the amino acid  
 40 sequence <SEQ ID 1462>. Analysis of this protein sequence reveals the following:

Possible site: 49

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4185(Affirmative) &lt; succ&gt;

bacterial membrane --- Certainty=0.0000(Not Clear) &lt; succ&gt;

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

An alignment of the GAS and GBS proteins is shown below:

Identities = 88/112 (78%), Positives = 102/112 (90%)

50          Query: 28   KLRADKKGTGTHRVAFENKRRLLKTAHLGICGRFVDKSLKYPHPLSAADHIVPIAKGGH 87  
               +LRADKKGTGTHRVAF++NR++LLK A +CSIOG+PVDKSLKYPHPLSAADHIVPIAKGGH  
               Sbjct: 3   QLRADKKGTGTHRVAFENKRRLLKTAHTVGOICGKFPVDKSLKYPHPLSAADHIVPIAKGGH 62

55          Query: 88   PSSIDNLQLTHWQCNQKSDKLFQNTAVRATVVGNNRLPQSRDWSYSASKE 139  
               PS+++NLQLTHWQCNQKSDKLF NQ +    +GNRNLQPSRDWSB+A K+  
               Sbjct: 63   PSALENLQLTHWQCNQKSDKLFQNTAVRATVVGNNRLPQSRDWSYAPKK 114

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 454

- 5 A DNA sequence (GBSx0491) was identified in *S. agalactiae* <SEQ ID 1463> which encodes the amino acid sequence <SEQ ID 1464>. Analysis of this protein sequence reveals the following:

Possible site: 33  
>>> Seems to have no N-terminal signal sequence

- 10 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.4481(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 455

A DNA sequence (GBSx0492) was identified in *S. agalactiae* <SEQ ID 1465> which encodes the amino acid sequence <SEQ ID 1466>. Analysis of this protein sequence reveals the following:

Possible site: 28  
>>> Seems to have no N-terminal signal sequence

- 20 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.2907(Affirmative) < succ>  
25 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

- 30 >GP:AAF43508 GB:AF145054 ORF15 [Streptococcus thermophilus  
bacteriophage 7201]  
Identities = 61/187 (32%), Positives = 90/187 (47%), Gaps = 31/187 (16%)  
Query: 1 MNTIEAKKLIDKOSIGKGGVGDIPVVKTHIVKVLDDIDQPOPEVPRFVADNYSKHKDSL 60  
NN +EA K I K+ + + L D I +P VP++VADNYS+HKD  
35 Sbjct: 1 MNRDEAVKKIAKEY-----ISIEHAEDLYDSIIT-KPVVPQYVADNYSBKDEF 49  
Query: 61 ECDL-----YLHMSIY--DREVEKDDPYVMQTSKRPVYTLINHQPGYTTQKEKLYT 112  
+L + H+ Y +E DF W +KN + L+NMHQPGY +KSK YT  
40 Sbjct: 50 YLNLHRVVRDFFHNLNAYTFNENPIDYDFACWYNTKQAIQILVNMHQPGYEVKKEKRYT 109  
Query: 113 VEIPN--FNERQLSFVLMRQLSCNVSIKVMHRNDLILKTDNDLQLTSEIRKDFDWAQ 170  
V I N E L+ R+ + RDN D +T + T E+ ++ + W  
Sbjct: 110 VRIRNLDEKETYNIDKFRS----TWVPYSRNDTDFRTIH---THKEL-ESGGPGWV 159  
45 Query: 171 FREBVE 177  
F E +E  
Sbjct: 160 FDCGIE 166

- A related GBS nucleic acid sequence <SEQ ID 10927> which encodes amino acid sequence <SEQ ID 10928> was also identified.

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1467> which encodes the amino acid sequence <SEQ ID 1468>. Analysis of this protein sequence reveals the following:

Possible site: 21

-556-

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

5           bacterial cytoplasm --- Certainty=0.3815 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 70/180 (38%), Positives = 98/180 (53%), Gaps = 30/180 (16%)

10   Query: 1   MNIESACKLIDKQSI-GKGGVGSDIPVVKTHIVKVLDDQIDQPQPEVPFVADWYERKHKDS 59  
           MNIEKAK+L+D   GK           V+K   V+ ++DQ++QP+PEVP+ VADW E+ K+  
   Sbjct: 1   MNIEKAKLVDNSKFYGKTS-----SVIKAE-VRDIIQQLNQPKPEVPQCVADWIERCKEE 55

15   Query: 60   LECDLYLYHMSIYDEEVEKKDDFYNNMQTSKNPVTLLNMHQFGYTIQEKLYTVIEIPN-- 117  
           DL L   ++           + W+ S   +           GYT++FEKLYTV++PN  
   Sbjct: 56   ---DLTL---KGLFNSNDMPAKIFDWIPGSDENCRLMARAWINGYTVFEKLYTVDLPNQG 110

20   Query: 118 PNERQLSFVLMRQLSGNVSIKVMHRDNLDLLKTLDDNDLQLTESEIRKDFDWAQCFREEVVE 177  
           P R ++ + Q                           L T+N ++L+TESEIRKDF+KAWQF EEV E  
   Sbjct: 111 PLVRGINTLYFSQN-----LATEN-VKLTESEIRKDFEWAQCFABEVTE 153

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 456**

A DNA sequence (GBSx0493) was identified in *S. agalactiae* <SEQ ID 1469> which encodes the amino acid sequence <SEQ ID 1470>. Analysis of this protein sequence reveals the following:

Possible site: 46

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

30           bacterial cytoplasm --- Certainty=0.5365 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

35

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 457**

A DNA sequence (GBSx0494) was identified in *S. agalactiae* <SEQ ID 1471> which encodes the amino acid sequence <SEQ ID 1472>. Analysis of this protein sequence reveals the following:

Possible site: 50

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

45   INTEGRAL   Likelihood = -8.55   Transmembrane   34 - 50 ( 31 - 54)

----- Final Results -----

50           bacterial membrane --- Certainty=0.4418 (Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9657> which encodes amino acid sequence <SEQ ID 9658> was also identified.

-557-

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1473> which encodes the amino acid sequence <SEQ ID 1474>. Analysis of this protein sequence reveals the following:

```

Possible site: 40
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood -11.25 Transmembrane 26 - 42 (20 - 49)

----- Final Results -----
bacterial membrane --- Certainty=0.5501(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 56/89 (62%), Positives = 71/89 (78%)

Query: 8 MTEQQMIDCLLYELAKKDKLNIRRNIIITPLSIVLAISILANVALQDHYKSQITLRTQL 67
 MTE+QMIDCLLYEL KKKK +++ II L+++L+ +S L V+L+ +Y+ QI LRTQL
Sbjct: 1 MTEBQMIDCLLYELVKKDKAIKKKSIITIAALTVMNLIVVSGLCVLSKYYEPQIYGLRTQL 60

Query: 68 SRTQKQLKRASDDRAQTKRIASLTNGG 96
 SRTQKQLKRAS+ RQTKRIA+LT NGG
Sbjct: 61 SRTQKQLKRASBQNRQTKRIADLTNNGG 89

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 458

A DNA sequence (GBSx0495) was identified in *S.agalactiae* <SEQ ID 1475> which encodes the amino acid sequence <SEQ ID 1476>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2040(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 459

A DNA sequence (GBSx0496) was identified in *S.agalactiae* <SEQ ID 1477> which encodes the amino acid sequence <SEQ ID 1478>. Analysis of this protein sequence reveals the following:

```

Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3044(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

```

-558-

bacterial outside --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

The protein has homology with the following sequences in the GENPEPT database:

5 >GP:AAD37108 GB:AF109674 unknown [Bacteriophage Tuc2009]  
 Identities = 50/143 (34%), Positives = 67/143 (45%), Gaps = 29/143 (20%)

Query: 1 MIPNFRFNFKETKGM-YG-VDGPELSVRKIYRCSLADDEFRCGRLETHPHFVEDNFDYIL 58  
 MIP RA++K+ ++M YG V+ F+ S+ YR HF +D  
 10 Sbjct: 1 MIPKLRAMDKQDERMSYGEVEYFDDSDIN--YRPD-----HPCTGADEDFVEF 44

Query: 59 MQSTGMFDRKNGVEIFDGDIVLTIRL-----IDY-TYKNFKGVKRMLEGRWLIDTGKDA 110  
 MQSTG+ DRKNGVEI++GDI+ + I Y Y G + EG L +  
 15 Sbjct: 45 MQSTGIKDRKNGVEIYGDIKLMHAIPLAPDCKIYGLYSPKYGYSIICEGNRLY---RQR 101

Query: 111 VGLWTEVDENEAIIGNIYQNSSELL 133  
 T E IGNIY+N ELL  
 20 Sbjct: 102 YWASTNKLNYEVIIGNIYENPELL 124

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1479> which encodes the amino acid sequence <SEQ ID 1480>. Analysis of this protein sequence reveals the following:

Possible site: 34  
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.4779 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

30 Identities = 44/52 (84%), Positives = 47/52 (89%)

Query: 1 MIPNFRFNFKETKGMVGDGPELSVRKIYRCSLADDEFRCGRLETHPHFVEDN 52  
 MIPNFR FNK+TKMY +DGF+ S RKIYRCSLADDEFRCGRLETHPHFVEDN  
 35 Sbjct: 1 MIPNFRGFNKKTKMYSIDGPKSSERKIYRCSLADDEFRCGRLETHPHFVEDN 52

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 240**

40 A DNA sequence (GBSx0497) was identified in *S.galactiae* <SEQ ID 1481> which encodes the amino acid sequence <SEQ ID 1482>. Analysis of this protein sequence reveals the following:

Possible site: 57  
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3843 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

50 A related GBS nucleic acid sequence <SEQ ID 9655> which encodes amino acid sequence <SEQ ID 9656> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.



Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 461

A DNA sequence (GBSx0498) was identified in *S.agalactiae* <SEQ ID 1483> which encodes the amino acid sequence <SEQ ID 1484>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.5189 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9653> which encodes amino acid sequence <SEQ ID 9654> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF43503 GB:AF145054 ORF10 [Streptococcus thermophilus
bacteriophage 7201]
Identities = 92/147 (62%), Positives = 121/147 (81%)

Query: 15 IEPKPCITRPFKPGTYEDPKMKWRKEVSGWIEKNYDGPFFDDCKIKVEVTFYMKAPKTL 74
IEPKPCITRPFKPGTYEDPKMKWR+E S IE+ YDG FF I V+VTFYMKAP ++
Sbjct: 7 IEPKPCITRPFKPGTYEDPKMKWRKECSRLIEQYDGGCFYGPISVDVTFYMKAPLSV 66

Query: 75 SKEPTORSKGKTIQVQNFVRELIVHAKKPDIDNLIAKAVFDSISDAGYDRIQSGIIVSD 134
SK+PT ++ KT ++ F+ E +VH++KPDIDNLIAK+VDSIS AGY+++ K GIVN+D
Sbjct: 67 SKKPTPKARAKTWDAFKKFAELHMSRKPDIDNLIAKALFDSISDAGYKRVKKGIVWTD 126

Query: 135 DNIVCDLRKKIKYQNPRIKVRIERID 161
D+IVC L A+K+YS+NPRI+ I+E++
Sbjct: 127 DSVICKLSAQKRYSENPRIRFPEIKELS 153
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 462

A DNA sequence (GBSx0499) was identified in *S.agalactiae* <SEQ ID 1485> which encodes the amino acid sequence <SEQ ID 1486>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4007 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-560-

**Example 463**

A DNA sequence (GBSx0500) was identified in *S.agalactiae* <SEQ ID 1487> which encodes the amino acid sequence <SEQ ID 1488>. This protein is predicted to be pXO1-07. Analysis of this protein sequence reveals the following:

```

5 Possible site: 26
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.3664 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15 >GP:AAC38715 GB:AF030367 maturase-related protein [Streptococcus pneumoniae]
 Identities = 146/373 (39%), Positives = 216/373 (57%), Gaps = 18/373 (4%)

 Query: 35 LYDRVYRKDKLVKVFYKRNKGSAGIDDFITIEIYAYGVQKFLDEIEDQLRNKKYQPKA 94
 L DK+ ++ + A+ VK NKGSAID TIEE+ Y Q + ++ ++ KY+P+
 Sbjct: 4 LLDKILSRERMLEAYNQVKSNGKSAGIDGMTIEEMDNYLRQNR-LTKELKQRKYQKP 62

20 Query: 95 VKRVYIPKANGKKRPLQIPTVRDRVVQTVAKIVIEIPFADFQEFVSYGFRPKRANQAIR 154
 V +V LPK +G R LGIP TV DR++Q A+ V+ PI E F + SYGFRP RS +AI
 Sbjct: 63 VLKVEIPKPGGIRQLGIP TVDRMIQQAIVQVMSPICEPHFSDTSYGFPRNRSKAM 122

25 Query: 155 BIYKYLNYGCEWVIDADLKGYFDTIPHDKLLLVKERVTDKSIKLLSLWLEAGIMEDNQ 214
 ++ +YLN G EW++D DL+ +FDT+P D+L+ LV + D L+ +L +G++ + Q
 Sbjct: 123 KLELYLNDGYEMIVDIDLEKFFDTVPQDRMSLVHNIIEDGFTESLIRKYLHSGVIINQ 182

30 Query: 215 VRSEILGTPQGQGVISPLANIYINALDRYKQNRLBORGHDAHLIRYADFVI-LCSINP 273
 ++GTPQGG +SPL+NI LN LD+ LE RG +RYADD VI + S
 Sbjct: 183 RYKTLGTPQGQGVISPLANIMLNELDK----ELEKRG--LRFVRYADDCVTTVGSAA 235

 Query: 274 KKYFYQYAKQRI--DKLGLITNEEKTRIVHATEGDFLGYTLRKSKSHKSKYKTYYPSPR 331
 K Y+ R +LGL +N KT+I E +LG+ KS + P +
35 Sbjct: 236 AKRVNYSVSRFTEKRGLKVMYTKTKITRPRE-LKYLQGFQWSSDGWGR-----PHQ 288

 Query: 332 KMSKSIKGVKVDVIQTGQHLLNPDMERLNPMLRGWANYFKAGNSKQHFSTINFYIYKL 391
 S++ K K+K + Q ++L +E+LN +RGW NYF GN K STD + L
40 Sbjct: 289 DSVRFKLLKLLTQKNSIDLTRIEQLNLSIRGWINTFSLGNKSTIVASIDERLTRKL 348

 Query: 392 TIMLAKEHKHSGK 404
 +++ K+ KK +
40 Sbjct: 349 RMIWKQMKKSR 361

```

45 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 464**

50 A DNA sequence (GBSx0501) was identified in *S.agalactiae* <SEQ ID 1489> which encodes the amino acid sequence <SEQ ID 1490>. Analysis of this protein sequence reveals the following:

```

 Possible site: 27
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
55 bacterial cytoplasm --- Certainty=0.3833 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9651> which encodes amino acid sequence <SEQ ID 9652> was also identified.

A further related DNA sequence (GBSx2517) was identified in *S.agalactiae* <SEQ ID 7217> which encodes the amino acid sequence <SEQ ID 7218>. Analysis of this protein sequence reveals the following:

```

Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3833 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1491> which encodes the amino acid sequence <SEQ ID 1492>. Analysis of this protein sequence reveals the following:

```

Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2299 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 113/163 (69%), Positives = 128/163 (78%), Gaps = 25/163 (15%)

Query: 1 MINNVLVGRMTKDLELRYTPSNQAVATFSLAVNRNRFNQSGEREADFNCVWIRQPAEN 60
 MINNVLVGRMTKDLELRYTPS AVATF+LAVNR FK+Q+GEREADFNCVWIRQ AEN
Sbjct: 1 MINNVLVGRMTKDLELRYTPSNQAVATFILA VNRITFSQNGEREADFNCVWIRQPAEN 60

Query: 61 LANWAKKGALVGITGRIQTRNYENQQGQRIYVTEVVAENFOLLESRSNSQ-----Q 111
 LANWAKKGAL+G+TGRIQTRNYENQQGQRIYVTEVVA+NFQ+LESR +++
Sbjct: 61 LANWAKKGALIGVTGRIQTRNYENQQGQRIYVTEVVAENFOLLESRATREGSGTSGPNNG 120

Query: 112 TNQSGNSNSY-----FGNANKMDISDDLFP 138
 N + +SSNSY FGN+N MDISDDLFP
Sbjct: 121 FNNRTSSNSYSAPAQOTFNFGRDDSPFGNSNFMNDISDDLFP 163

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 465

A DNA sequence (GBSx0502) was identified in *S.agalactiae* <SEQ ID 1493> which encodes the amino acid sequence <SEQ ID 1494>. Analysis of this protein sequence reveals the following:

```

Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -1.33 Transmembrane 17 - 33 (17 - 33)

----- Final Results -----
bacterial membrane --- Certainty=0.1532 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 466

A DNA sequence (GBSx0503) was identified in *S. agalactiae* <SEQ ID 1495> which encodes the amino acid sequence <SEQ ID 1496>. This protein is predicted to be p22 crf-like protein. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence
```

----- Final Results -----

```
bacterial cytoplasm --- Certainty=0.2469(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAA97824 GB:AB044554 orf 17 [Staphylococcus aureus prophage  
phiPV83]  
Identities = 93/183 (50%), Positives = 120/183 (64%), Gaps = 5/183 (2%)

Query: 1 MRKSESITEYAKAFCKAQLEVKQPLKDKDNPFKSKYVPLENVTEAITTAFANNGISFSQ 60  
M KSB++ E KA + EVKQPLKDK+NPFFKSKYVPLENV EAI A +G+S++Q  
Subject: 1 MNKSETVVEINKAMVAFRKEVKQPLKDKDNPFKSKYVPLENVVRAIDEAATPHGLSYTQ 60

Query: 61 DPTTNTENG YIDVATLMMHTSGEWEVEYGLPSVKPTKNDVQAGSAITYAKRYALSAIFGI 120  
N +G + VAT++NH SGE++EY P+ + KN QGAGS I+Y KRY+LSAIFGI  
Subject: 61 W-ALNDVDGRVG VATLMMHESGEYIEYDPVFMNAEKNTDQAGSLISYLKRYSLSAIFGI 119

Query: 121 TSDQDDGGNEDSKPNNRQSPKATTKKTQRTGYQTPKISNIQIETYSKDLNDIAKATNQ 180  
TSDQDDGGNE S MN +PK T +TQ +T I ++ ++ + K QN  
Subject: 120 TSDQDDGGNEASGKIN---NPKCOT-RTQWASSETIGLRKEVISFTKLKGTDKPEAP 175

Query: 181 VEE 183  
+ E  
Sbjct: 176 IVE 178

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 467

A DNA sequence (GBSx0504) was identified in *S. agalactiae* <SEQ ID 1497> which encodes the amino acid sequence <SEQ ID 1498>. This protein is predicted to be gp157. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence
```

----- Final Results -----

```
bacterial cytoplasm --- Certainty=0.3148(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD44102 GB:AF115103 orf157 gp [Streptococcus thermophilus  
bacteriophage sf121]

-563-

Identities = 59/160 (36%), Positives = 100/160 (61%), Gaps = 3/160 (1%)

Query: 1 MAYLYELGSIYALQSQMDLDETFQDTLOSIDFQSDLENNIYFVFMIGNVQADAKYKA 60  
 MA LYSL G + + +M++D+ET DTL+ID+ SD EN +E +VK++K++AD E K  
 5 Sbjct: 1 NATLYELTGQFLFYNNMEIDDETGLDTLEALDWTSDYENKVBGVYKVIKSLADTEARSN 60

Query: 61 EKRAFYKKQQAARAKAKYKRTIRLAMELSQKKQVDAGMFVSIARRSKKVELDETCTPL 120  
 EK+ K ++K +K K + ++M + + +VD +FK+ +SK V ++E K+P  
 10 Sbjct: 61 EKKRLDGLNKSQSKICIDKLGARLAISMTEGQTRVDITLFCIGPHKSKAV-VVMKELPK 119

Query: 121 DYMCEKIKYKPMKARISKALKSGIDISGVLEIKTSSLOVK 160  
 +Y + YKP K + + LKSG I G L E +L ++  
 Sbjct: 120 EY--QIATYKPKKKTLELLKSGKHIGATLEERRNLNIR 157

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 468

A DNA sequence (GBSx0505) was identified in *S.agalactiae* <SEQ ID 1499> which encodes the amino acid sequence <SEQ ID 1500>. This protein is predicted to be tropomyosin 2. Analysis of this protein sequence reveals the following:

Possible site: 26  
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.4474 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

30 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 469

35 A DNA sequence (GBSx0506) was identified in *S.agalactiae* <SEQ ID 1501> which encodes the amino acid sequence <SEQ ID 1502>. Analysis of this protein sequence reveals the following:

Possible site: 21  
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.4114 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

45 A related GBS nucleic acid sequence <SEQ ID 9649> which encodes amino acid sequence <SEQ ID 9650> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 470

A DNA sequence (GBSx0507) was identified in *S.agalactiae* <SEQ ID 1503> which encodes the amino acid sequence <SEQ ID 1504>. Analysis of this protein sequence reveals the following:

Possible site: 17  
>>> Seems to have no N-terminal signal sequence

#### ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3799 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1505> which encodes the amino acid sequence <SEQ ID 1506>. Analysis of this protein sequence reveals the following:

Possible site: 31  
>>> Seems to have no N-terminal signal sequence

#### ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3775 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 43/46 (93%), Positives = 46/46 (99%)

Query: 1 NTRQHRETLIWRASHQERERLLDPGLVDKSYVTLRLQRLKKYAI 46  
NTRQHRETLIWRASHQERERLLDPGLVDKSYVTLRLQRLKKYAI  
Sbjct: 1 NTRQHRETLIWRASHQERERLLDPGLVDKSYVTLRLQRLKKYAI 46

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 471

A DNA sequence (GBSx0508) was identified in *S.agalactiae* <SEQ ID 1507> which encodes the amino acid sequence <SEQ ID 1508>. Analysis of this protein sequence reveals the following:

Possible site: 61  
>>> Seems to have no N-terminal signal sequence

#### ----- Final Results -----

bacterial cytoplasm --- Certainty=0.4306 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1509> which encodes the amino acid sequence <SEQ ID 1510>. Analysis of this protein sequence reveals the following:

Possible site: 61  
>>> Seems to have no N-terminal signal sequence

#### ----- Final Results -----

-565-

bacterial cytoplasm --- Certainty=0.4308(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below:

Identities = 76/77 (98%), Positives = 76/77 (98%)

Query: 1 MDQSIFFNFNKKIKD FGKTASKETPAKFASYCABGIEKNGVKPIFNWNLINLYAFQTGMT 60

MDQSIFFNFNKKIKD FGKTASKETPAKFASYCABGIEKNGVKPIFNWNLINLYAFQTGMT

10 Sbjct: 1 MDQSIFFNFNKKIKD FGKTASKETPAKFASYCABGIEKNGVKPIFNWNLINLYAFQTGMT 60

Query: 61 AEADRLRIERYKQENTL 77

AEADRLRIERYKQEN L

15 Sbjct: 61 AEADRLRIERYKQENAL 77

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 472

A DNA sequence (GBSx0509) was identified in *S.agalactiae* <SEQ ID 1511> which encodes the amino acid sequence <SEQ ID 1512>. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2706(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1513> which encodes the amino acid sequence <SEQ ID 1514>. Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3316(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 52/127 (40%), Positives = 75/127 (58%), Gaps = 1/127 (0%)

Query: 160 EDRPFDVVEAMLCRLGKVKFEDMINDVILIGQNVSKDLFLEAVKVAVANNVRKFNITARIL 219

E + + + GR + FE + I ++ N+ ++ A++ AV NN + YI +IL

45 Sbjct: 3 EKKLPEFPQLTFGRMISPPELIDIQNIHEDNMPIEVNIALREAVERNKISWKYINKIL 62

Query: 220 INMINDGIKTPEQAVQAQDFKAKKANKTQSQSNVPSNSNPDYKGEPLKEFALGSIDDI 279

+W G T E+ + F K +++ +SNVPSNSNPDYK PDL+EFALGS+D I

Sbjct: 63 VDWYKSGDITVEKVRDLRQRDSSKKQSVIT--SNVPSNSNPDYKPEPLKEFALGSMDGI 121

Query: 280 EDGSGDF 286

EDGSGDF

50 Sbjct: 122 EDGSGDF 128

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 473**

A DNA sequence (GBSx0510) was identified in *S.agalactiae* <SEQ ID 1515> which encodes the amino acid sequence <SEQ ID 1516>. Analysis of this protein sequence reveals the following:

5       Possible site: 26  
       >>> Seems to have an uncleavable N-term signal seq  
       INTEGRAL   Likelihood = -5.63   Transmembrane   13 - 29 ( 11 - 31)  
  
       ----- Final Results -----  
       bacterial membrane --- Certainty=0.3251(Affirmative) < succ>  
 10       bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
       bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9647> which encodes amino acid sequence <SEQ ID 9648> was also identified.

15   The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 474**

20   A DNA sequence (GBSx0511) was identified in *S.agalactiae* <SEQ ID 1517> which encodes the amino acid sequence <SEQ ID 1518>. Analysis of this protein sequence reveals the following:

      Possible site: 34  
       >>> Seems to have no N-terminal signal sequence  
  
 25       ----- Final Results -----  
       bacterial cytoplasm --- Certainty=0.5822(Affirmative) < succ>  
       bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
       bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30   The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 475**

35   A DNA sequence (GBSx0512) was identified in *S.agalactiae* <SEQ ID 1519> which encodes the amino acid sequence <SEQ ID 1520>. Analysis of this protein sequence reveals the following:

      Possible site: 13  
       >>> Seems to have no N-terminal signal sequence  
  
 40       ----- Final Results -----  
       bacterial cytoplasm --- Certainty=0.4175(Affirmative) < succ>  
       bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
       bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45   The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.



-567-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 476

A DNA sequence (GBSx0513) was identified in *S.agalactiae* <SEQ ID 1521> which encodes the amino acid sequence <SEQ ID 1522>. This protein is predicted to be P1-antirepressor homolog. Analysis of this protein sequence reveals the following:

Possible site: 37  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3411(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9645> which encodes amino acid sequence <SEQ ID 9646> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAG31333 GB:AF182207 ORF 169a [Bacteriophage mv4]  
Identities = 88/167 (52%), Positives = 122/167 (72%)

Query: 100 MLQRNEKSKQVKRYFIQVEKDFNSPEKIMARALLNADKKITMLTWNQQLQDLKKAQK 159  
M+ + K K+ + R+ YFIQVEK+ + NSPE I+ RAL + + + + I L + N L L L+ E+ K+  
Sbjct: 1 MMSKTARGKETQYFIQVEKNNSPFEMIIQALEISNARIQELQAQKSLITLQESNNK 60

Query: 160 ARYLDLIIESKGLRVTQIAADYQMSVKNPKNTLLRFGVQKRVNGQWILYKRMHGGTYD 219  
A YLD+I+ + L TQIAADYQ S FN+ L E G+CHKVNGQWILYK +MGGTY  
Sbjct: 61 ASYLDIILGTPELLATTQIAADYGSARTFNQLLKEVGIGKRVNGQWILYKAYMGGTYQ 120

Query: 220 SHTFDYQDKNGHTRANVTITWTQKGRPLVYELLKDNNILFLIEQEDI 266  
S +F +D+ GH R+ +T WTQKGR +Y+LK+N LFLIE+DI  
Sbjct: 121 SKSFAPKDRNGHDRGKPFSTYWTQKGRKLIYDVLKENTLFLIERDDI 167

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1523> which encodes the amino acid sequence <SEQ ID 1524>. Analysis of this protein sequence reveals the following:

Possible site: 19  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4214(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 130/249 (52%), Positives = 163/249 (65%), Gaps = 14/249 (5%)

Query: 19 MNQLINITLWNGQEPVVSGRDLHNVNLNKTQYTKWLKRMSEYGFENVDYIALSQKRITA 78  
MNQLIN+TLWNGQEPVVSGRDLH VL IKQYTKWLKRMSEYGF EN D+ +AISQRUITA  
Sbjct: 1 MNQLINVTLWNGQEPVVSGRDLHKVLEIKQYTKWLKRMSEYGFENEDFMALISQRUITA 60

Query: 79 QGNRTYIDHVLKLMARKEIAMLQRNEKSKQVKRYFIQVEKDFNSPEKIMARALLNADKK 138  
QGN+TEY DHVLKLMARKEIAMLQRNEKSK+VKRYFIQVEKDFNSPEKIMARALLNADKK  
Sbjct: 61 QGNQTEYIDHVLKLMARKEIAMLQRNEKSKQVKRYFIQVEKDFNSPEKIMARALLNADKK 120

Query: 139 ITNLTMENQQLQLDLKKAQKQARYLDLITBSKGLRVTQIAA-----DYGMSVKNPKNTLL 193  
+ +L+ + + + + D+ S + + V +A + + + L  
Sbjct: 121 V-----HKLKQIETADRPKVLPAADAVSASHSTILVGLAKLILKQNGVNTIGATLFTWL 173

-568-

Query: 194 LEPGVQHKVAGQ-WIL-YKRIMGKGYTDSHTPDYQDKRGHTRANVTITWTQKGRLEFLYLE 251  
 + G K NG+ W + ++ + G +GH + T T KG+ +  
 Sbjct: 174 RKHGYLEIKRGRDWMPTQKSVELGLIRKVSITSHDGHITVSKTFLVTGKQQQFINK 233

5 Query: 252 LKDNINILPL 260  
 + LP+  
 Sbjct: 234 FLNQSEYLPV 242

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 477

A DNA sequence (GBSx0514) was identified in *S. agalactiae* <SEQ ID 1525> which encodes the amino acid sequence <SEQ ID 1526>. Analysis of this protein sequence reveals the following:

Possible site: 44  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.4205 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1527> which encodes the amino acid sequence <SEQ ID 1528>. Analysis of this protein sequence reveals the following:

Possible site: 32  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 21/63 (33%), Positives = 31/63 (49%), Gaps = 1/63 (1%)

Query: 1 MQQPNLKQLREKKGPTQWELADKANUSRLVUGLETGYSYSETSTASLKKLAKALDVKID 60  
 M+ LK R K +Q LAD VSR + +E G Y+ T + + + LD + D  
 Sbjct: 1 MNLKLKAARAGKDLQQALADLVGVSRQTIAAVEKGDYNTFNLICI-AICRVLDKTLDD 59

40 Query: 61 LFF 63  
 LF+  
 Sbjct: 60 LFW 62

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 478

A DNA sequence (GBSx0515) was identified in *S. agalactiae* <SEQ ID 1529> which encodes the amino acid sequence <SEQ ID 1530>. Analysis of this protein sequence reveals the following:

Possible site: 26  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.0396 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

-569-

bacterial outside --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

The protein has homology with the following sequences in the GENPEPT database:

5 >GP:BA17582 GB:D90907 hypothetical protein [Synechocystis sp.]  
 Identities = 45/164 (27%), Positives = 79/164 (47%), Gaps = 33/164 (20%)

Query: 102 KEELRLNLFKLIASSMDKSONEPNHPSPFTEIIRKQFKIDAQNFKIISDLYFKGPFVATG 161  
 ++E L+ L+AS++ +S + SF+E++KQ D +DA+ ++ L+ +  
 Sbjct: 97 DDENLQTLNANLLASALTESDRTNYSKSFVFLKQVDIVDAELNVLVYLHLRV----- 150

10 Query: 162 TYYTTIIGQDKPLEHIASHVFDVNLQNDIAQSSSLTNLERLGLIQTINY--KAHVDEK 219  
 KP E ++ D+ + N + I S +L NLERLGL+ I+ + VDE+  
 Sbjct: 151 -----MAKPDFTYAN---DSRKYNIVCI-SVALNNLERLGLLIHKKYDDTDFDEA 198

15 Query: 220 YYNIILNNSFITKNSBLKBNKRVLTNLGMITLTLFGVRFKTC 263  
 +I ++ N K ++LTFG+ F + C  
 Sbjct: 199 RISIW---YMQGNRSFKAH-----VSLTLFGIHPNRVC 229

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1531> which encodes the amino acid sequence <SEQ ID 1532>. Analysis of this protein sequence reveals the following:

20 Possible site: 19  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.0151 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

30 Identities = 64/215 (29%), Positives = 105/215 (48%), Gaps = 23/215 (10%)

Query: 65 QKLAKETIQVVSKNIE-NLQEPSLSTAGPALEASKFYLEEBELRLNLFKLIASSMDKSN 123  
 +K EI SK + +L+EP I PA+ S+ YL E LRH+P + IAS+ ++ K  
 Sbjct: 72 EKFNGEIDCEFSKIPQTLSEKPEVYILNPAINESECYLSNETLRNMFARTLASTFNQKE 131

35 Query: 124 EFNHPSPFTEIIRKQFKIDAQNFKIISDLYFKGPFVATGTYTTIIGQDKPLEHI----- 177  
 + H +F++IIRK +DAQN +I+ IG E++  
 Sbjct: 132 KDLHSAFVQIIKQMTPLDAQNILLIQ-----EGNNLIANLQIGVHYSEKNSGTVNK 184

40 Query: 178 ASHVFDVNLQNDIAQSSSLTNLERLGLIQTINYKAHVDEKRYTNILNNSFITKNSBLK 237  
 A+++++ L+ + + I +SS+ NL RLGLI++Y + + Y +I + SE+  
 Sbjct: 185 ANNTIYKSLDYSPDII-ASSIDNLTSLGLIKVDYLRPLDSNYESIKQTITYKSLSEIN 243

45 Query: 238 EQNKRVLTNL-----GMITLTLFGVRFKTC 264  
 N +N G ++LTFG +F CL  
 Sbjct: 244 TLNLPKTSNPKYDIKIKKGVSLIDPQKPFISVCL 278

SEQ ID 1530 (GBS261) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 8; MW 31kDa).

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 479**

A DNA sequence (GBS0516) was identified in *S. agalactiae* <SEQ ID 1533> which encodes the amino acid sequence <SEQ ID 1534>. Analysis of this protein sequence reveals the following:

55 Possible site: 16  
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -8.55 Transmembrane 3 - 19 ( 1 - 26)

-570-

## ----- Final Results -----

5           bacterial membrane --- Certainty=0.4418(Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 480**

A DNA sequence (GBSx0517) was identified in *S.agalactiae* <SEQ ID 1535> which encodes the amino acid sequence <SEQ ID 1536>. Analysis of this protein sequence reveals the following:

Possible site: 47

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

INTEGRAL   Likelihood = -4.99   Transmembrane   35 - 51 ( 30 - 51)

## ----- Final Results -----

20           bacterial membrane --- Certainty=0.2996(Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1537> which encodes the amino acid sequence <SEQ ID 1538>. Analysis of this protein sequence reveals the following:

Possible site: 47

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

INTEGRAL   Likelihood = -4.94   Transmembrane   31 - 47 ( 30 - 51)

## ----- Final Results -----

30           bacterial membrane --- Certainty=0.2975(Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

35 An alignment of the GAS and GBS proteins is shown below:

Identities = 45/52 (86%), Positives = 48/52 (91%)

Query: 1 MNNKKLMLGDLERTFTSRDQKERTSVFEGGVLPAALLVLGGITWLIWLTIK 52

MNNKKLM GDLERTFT+ DQKERTS+EFEGGVLPAALLVLGGI W+IWM ITK

40 Sbjct: 1 MNNKKLMFGDLERTFTNHDQKERTSIEFEGGVLPAALLVLGGIWNIAWFTIK 52

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 481**

45 A DNA sequence (GBSx0518) was identified in *S.agalactiae* <SEQ ID 1539> which encodes the amino acid sequence <SEQ ID 1540>. Analysis of this protein sequence reveals the following:

Possible site: 35

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

## ----- Final Results -----

50           bacterial cytoplasm --- Certainty=0.3445(Affirmative) < succ>

-571-

```

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

- 5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 482

- 10 A DNA sequence (GBSx0519) was identified in *S.agalactiae* <SEQ ID 1541> which encodes the amino acid sequence <SEQ ID 1542>. Analysis of this protein sequence reveals the following:

```

Possible site: 37
>>> Seems to have no N-terminal signal sequence

```

- 15 ----- Final Results -----
- ```

bacterial cytoplasm --- Certainty=0.3934 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

- 20 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 483

- 25 A DNA sequence (GBSx0520) was identified in *S.agalactiae* <SEQ ID 1543> which encodes the amino acid sequence <SEQ ID 1544>. This protein is predicted to be repressor protein. Analysis of this protein sequence reveals the following:

```

Possible site: 61
>>> Seems to have no N-terminal signal sequence

```

- 30 ----- Final Results -----
- ```

bacterial cytoplasm --- Certainty=0.0905 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

- 35 A related GBS nucleic acid sequence <SEQ ID 9643> which encodes amino acid sequence <SEQ ID 9644> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1545> which encodes the amino acid sequence <SEQ ID 1546>. Analysis of this protein sequence reveals the following:

- 40 Possible site: 55
- ```

>>> Seems to have no N-terminal signal sequence

```
- 45 ----- Final Results -----
- ```

bacterial cytoplasm --- Certainty=0.3117 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 175/264 (66%), Positives = 207/264 (78%), Gaps = 19/264 (7%)

```

-572-

Query: 34 LGKYIKYVETNNLSMAEFKESGTSKAY--VSILEKNRDPNGKRETIIPISPIIKKVSDT 91  
 LG I+K R+ N++ E++ G+ K Y VS EKN + GK++ KK+++  
 5 Sbjct: 24 LGDRIRKLEGRNMTQVELSELGM-KTYYTVSKNEKNENFPKGRDL-----KKLAET 75

Query: 92 IGISFDOLIASLNRNQIALNKTKTERNLTSSTLQKITSTSSQLBQPRQKVLSPANEQL 151  
 ++ D LL L++K K + +I S +QLBQPRQKVL+PANEQL  
 Sbjct: 76 FNVTSVLLG-----LTDSLGKITIQNBQPEIVSIYNQLBQPRQKVL+PANEQL 126

10 Query: 152 HEQNKVVMFDRKVERTENYITDYVEGLVAAGLGAYQEDNLHMEVTKLRADVDVPOKDYDTA 211  
 HEQNK VS+PD+K EETE+YITDYVEGLVAAGLGAYQEDNLHM+VKLR+DVPD+YDTA  
 Sbjct: 127 HEQNKTVSIFDKKSEETDYITDYVEGLVAAGLGAYQEDNLHMKVLRSDVDPDEYDTA 186

15 Query: 212 KVAGNSMEPLIQNDLLPVKVSQVNMNDIGIPQVNGKHFPVKLRDYDGAWYLSLNKS 271  
 KVAG+SMELPIQNDLLF+KVSSQVNMNDIGIPQVNGKHFPVKLRDYDGAWYLSLNKS  
 Sbjct: 187 KVAGDSMEPLIQNDLLPIKVSQVNMNDIGIPQVNGKHFPVKLRDYDGAWYLSLNKS 246

Query: 272 YEEIYLSENNIRITGEVVDIYRE 295  
 YEEIYLS++D+IRTIGEVDIYRE  
 20 Sbjct: 247 YEEIYLSKDDIRITIGEVDIYRE 270

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 484

- 25 A DNA sequence (GBSx0521) was identified in *S. agalactiae* <SEQ ID 1547> which encodes the amino acid sequence <SEQ ID 1548>. Analysis of this protein sequence reveals the following:

Possible site: 45  
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.3760 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 35 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 485

- 40 A DNA sequence (GBSx0522) was identified in *S. agalactiae* <SEQ ID 1549> which encodes the amino acid sequence <SEQ ID 1550>. This protein is predicted to be integrase (ripX). Analysis of this protein sequence reveals the following:

Possible site: 20  
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.2719 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 50 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB96616 GB:AJ400629 integrase [Streptococcus pneumoniae  
 bacteriophage M1]  
 Identities = 36/59 (61%), Positives = 48/59 (81%), Gaps = 1/59 (1%)

-573-

Query: 2 KLYGDYHTLFRHSHISFLAENGIPLANAIMDRVGHSDPKTTLSIYSHTTVMKKE-IINK 59  
 KI + +H+FRHSHISFLAS G+P+ +IMDRVGHSD+ K TL IYSHTT +M++ ++NK  
 Sbjct: 312 KIEKNLSHIFRSHISFLAENGIPKISIMDRVGHSDNAIMTLEYSHTTMEDKLVNK 370

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1551> which encodes the amino acid sequence <SEQ ID 1552>. Analysis of this protein sequence reveals the following:

Possible site: 20  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2719 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 63/71 (88%), Positives = 66/71 (92%)  
 Query: 1 MKIKGDYHTLFRHSHISFLAENGIPLANAIMDRVGHSDPKTTLSIYSHTTVMKKEIINK 60  
 +KIKGDYHTLFRHSHISFLAENGIPLANAIMDRVGHSDPKTTLSIYSHTTVMKKEIINK  
 Sbjct: 1 LKIKGDYHTLFRHSHISFLAENGIPLANAIMDRVGHSDPKTTLSIYSHTTVMKKEIINK 60  
 Query: 61 TAPFVPLKSE 71  
 T PF +K +  
 Sbjct: 61 TDPFKTQIKQK 71

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 486

A DNA sequence (GBSx0523) was identified in *S.agalactiae* <SEQ ID 1553> which encodes the amino acid sequence <SEQ ID 1554>. This protein is predicted to be 50S ribosomal protein L19 (rpL5). Analysis of this protein sequence reveals the following:

Possible site: 54  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3331 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9641> which encodes amino acid sequence <SEQ ID 9642> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC01534 GB:U88973 ribosomal protein L19 [Streptococcus thermophilus]  
 Identities = 110/115 (95%), Positives = 112/115 (96%)  
 Query: 25 MNPLIQSLTSEQLRSDIPEFRACGTVRVHAKVVEGTERRIQIFEGVVISRKGQGISSEMYT 84  
 MNPLIQSLTSEQLR+DIP FR GLTVRVHAKVVEGTERRIQIFEGVVISRKGQGISSEMYT  
 Sbjct: 1 MNPLIQSLTSEQLRSDIPEFRACGTVRVHAKVVEGTERRIQIFEGVVISRKGQGISSEMYT 60  
 Query: 85 VRKISGGIGVERTFPIHTPRVDKIEVVRYGKVRRAKLYLRAIQGAARIKER 139  
 VRKIS GIGVERTFPIHTPRVDKIEVVRYGKVRRAKLYLRAIQGAARIKER+  
 Sbjct: 61 VRKISSGIGVERTFPIHTPRVDKIEVVRYGKVRRAKLYLRAIQGAARIKER 115

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1555> which encodes the amino acid sequence <SEQ ID 1556>. Analysis of this protein sequence reveals the following:

```

Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4849 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 111/115 (96%), Positives = 113/115 (97%)

Query: 25 MNPLIQSLTEGQLRSDIPEFRAGDTVRVHAKVVEGTRERIQIFEGVVISRKQGISSEMYT 84
 MNPLIQSLTEGQLRSDIP FR GDTVRVHAKVVEGTRERIQIFEGVVISRKQGISSEMYT
Sbjct: 1 MNPLIQSLTEGQLRSDIPNFRPGDTVRVHAKVVEGTRERIQIFEGVVISRKQGISSEMYT 60

Query: 85 VRKISGGIGVERTFPPIHTPRVDKIEVVRKGVRRKALYYLRALQGGAARIKEIRR 139
 VRKISGGIGVERTFPPIHTPRVDKIEV+R-GKVRRAKLYLRALQGGAARIKEIRR
Sbjct: 61 VRKISGGIGVERTFPPIHTPRVDKIEVRHGKVRRAKLYLRALQGGAARIKEIRR 115

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 487

A DNA sequence (GBSx0524) was identified in *S.agalactiae* <SEQ ID 1557> which encodes the amino acid sequence <SEQ ID 1558>. This protein is predicted to be ISL2 protein. Analysis of this protein sequence reveals the following:

```

Possible site: 58
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:CAC18596 GB:AJ276419 IS1381 transposase [Streptococcus pneumoniae]
Identities = 111/129 (86%), Positives = 117/129 (90%)

Query: 1 MKAQAIVTSQGRIVSLDIANYCHDMKLPMSRRNIQQAAILADSGYQIGMYSQAQT 60
 MK QAIVTSQGRIVSLDI VNYCHDMKLPMSRRNIQQA KILADSGYQG+MK+Y QAGT
Sbjct: 1 MTKQAIVTSQGRIVSLDITVNYCHDMKLPMSRRNIQQAAILADSGYQGLMKIYQAQT 60

Query: 61 PRKSSKLKPLTLEDKTYNHTLSKERIKVENIFAKVKTPKIPSTTYRNRRRFGIARNLIA 120
 RKSSKIKPLT+EDK NH LSKRR KVENIFAKVKTPK+PSTTYR+ RRPGLRNL A
Sbjct: 61 BRKSSKLKPLTLEDKACHALSKEKSRVENIFAKVKTFMPSTTYRSHRFRGLRNLISA 120

Query: 121 GVINRELGP 129
 G+IN ELGP
Sbjct: 121 GIINRELGP 129

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.



-575-

**Example 488**

A DNA sequence (GBSx0526) was identified in *S. agalactiae* <SEQ ID 1559> which encodes the amino acid sequence <SEQ ID 1560>. Analysis of this protein sequence reveals the following:

- Possible site: 61  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -10.99 Transmembrane 81 - 97 ( 67 - 107)  
 INTEGRAL Likelihood = -6.32 Transmembrane 8 - 24 ( 6 - 25)  
 INTEGRAL Likelihood = -2.76 Transmembrane 120 - 136 ( 120 - 136)
- 10 ----- Final Results -----  
 bacterial membrane --- Certainty=0.5394 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
- 15 The protein has homology with the following sequences in the GENPEPT database:  
 >GP:BAB04382 GB:AF001509 unknown conserved protein in others  
 [Bacillus halodurans]  
 Identities = 53/150 (35%), Positives = 82/150 (54%), Gaps = 1/150 (0%)
- 20 Query: 1 MLNPYKRIPTLGLLATFLLFIHFGRYSGLGTNLIEASPTNKNLYDDYDWLLKLCI/VITL 60  
 M N R F GL+ L +I Y+G G +++E SPT +++ Y +L KL T +T+  
 Sbjct: 251 MKNHTVRAPVGGILITVALTYIIGSYDYNRGRLDLEDSPT-QDVPPYAFIAKLVPATVM 309
- Query: 61 AAGYQGGVETPLFAIGASLGVIIPILGLFVILVAALGYTSVFSQATNTLLGPILGGEV 120  
 G+ GGE PLF +GA+LG + + LP+ +AALG FG NT+ L+G E+  
 25 Sbjct: 310 GMGPFVGGELPLFFVGATLGNLTHAFIDLPFLAALGNIVTFGGGANTPIAAFLGVEM 369
- Query: 121 FGFANTPYFVIIVCLVAYSISHAHTIYGAQS 150  
 F +F + CL +Y S H ++ O+  
 30 Sbjct: 370 FNGKGEIYFFVACLTSYLFSGHKGMLPQST 399

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1561> which encodes the amino acid sequence <SEQ ID 1562>. Analysis of this protein sequence reveals the following:

- Possible site: 35  
 >>> Seems to have no N-terminal signal sequence
- 35 INTEGRAL Likelihood = -11.99 Transmembrane 56 - 72 ( 53 - 78)  
 INTEGRAL Likelihood = -7.17 Transmembrane 337 - 353 ( 327 - 355)  
 INTEGRAL Likelihood = -6.74 Transmembrane 264 - 280 ( 260 - 282)  
 INTEGRAL Likelihood = -6.16 Transmembrane 167 - 183 ( 161 - 187)  
 40 INTEGRAL Likelihood = -5.26 Transmembrane 223 - 239 ( 217 - 242)  
 INTEGRAL Likelihood = -5.10 Transmembrane 20 - 36 ( 19 - 42)  
 INTEGRAL Likelihood = -0.37 Transmembrane 102 - 118 ( 102 - 119)  
 INTEGRAL Likelihood = -0.16 Transmembrane 300 - 316 ( 300 - 316)
- 45 ----- Final Results -----  
 bacterial membrane --- Certainty=0.5798 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
- 50 The protein has homology with the following sequences in the databases:  
 >GP:BAB04382 GB:AF001509 unknown conserved protein in others  
 [Bacillus halodurans]  
 Identities = 129/397 (32%), Positives = 210/397 (52%), Gaps = 14/397 (3%)
- 55 Query: 20 VLGLVGLALPIGGAGVVVDVIFGKGLLPISRYRDHHLFLLPFLAAGLVLVFLXKLG- 78  
 +L + + IG VG + L E R++ + +L FL LAQL + +LY K G  
 Sbjct: 9 LILTWIFFGHIMIGAVSAGLALLTVNDHLARTRENKRPWPL-PLPLAGLALGYLYMKAGT 67
- Query: 79 ---KEVRQGMOLVFQVGHGQKNQIPMLILPLLSTVTVTHLFQASAGRGVAVQIGATIS 135  
 E+ +G LV + G K ++ L PL+ T++T LEF S GREG A+Q+G +++  
 60 Sbjct: 68 SAGNELYKGNLVIERSVQG-KKKMLLRIGPLVVLGTPTMTLFGGSGTREGGAALQMGGSA 126

-576-

Query: 136 HYCRR-FVTSQBAARHLLIMGAAGFAGLQPTPIAAVFALEVLVGTIRYSALLPLVA 194  
 + F R LL+ G++G+ F TPI A +F +E+ +G L++ AL+P DNA  
 Sbjct: 127 EAVNLKFKVLIDTRILLMGISGAGGAAGPTITAAIFGMBASLRLKFEALVPCINA 186

Query: 195 AYVASWTSIALG-LSKFTVILETHITITPLIVKLIGLIFGLVGNFPAYLE-GWPKFY 252  
 ++V +T+ +E ++ ++ LT K+I L ++F LV + L G K  
 Sbjct: 187 SFVGHYTTKFMHIVSHEKFIATVFEVSALTPSKVILLATVPSLVSVLTCQLRHSIHKL 246

Query: 253 LSKLENPFYRIAPFGALLSICL--HIGHVGRYSGLTNLIAAAPSQQTILTYDWLLKMI 310  
 + N R AF+G L+ + L +IG Y+G G +++ +F+ Q + Y +L K++  
 Sbjct: 247 EKYTMKNTVTR-AFVGILLIIVALTITIGSTD-YNGRGLMDLEDST-QDVPPYAPLAKLV 303

Query: 311 VTVISLSAGFOGSEVTPFLFAIGASLGIVLAPYLGLPVLVAALGYTTVFGSATNTFWAFI 370  
 T +++ GF GGE PLF +GA+IG L ++ LP+ +AALG FG NP A  
 Sbjct: 304 FTAVTMGMGFGVSGEAPLFFVGATIGTLHAPLIDPLSLAALGMIVTFGGANTPIPAF 363

Query: 371 FIGIEVFGPENALAYPVTSAANVMVSHRHSITYQK 407  
 +G+R+ + + +FV +Y+ S H ++ Q +  
 Sbjct: 364 LLGVEMFKGKIEFFVACLNTYLFSGHHGLMPQSTI 400

An alignment of the GAS and GBS proteins is shown below:

Identities = 91/147 (61%), Positives = 111/147 (74%)

Query: 3 NPYRIFTIGLLATFLFIFHFGYSGGLTNLIEASFTNNKLYDYDMLKLCITVITLAA 62  
 NPY RI +G L + L I H GRYSGGLTNL L A+P+ + + YDWLKR+ +TVI+L+A  
 Sbjct: 259 NPYRIAFIIGALLSICIMIGHVGRYSGLGNLIAAAPSQQTILTYDWLLKMTVTVISLEA 318

Query: 63 GYQGGSEVTPFLFAIGASLGVIIAPILGLFVLVAALGYTTSVFGSATNTLLGPILGSEVFG 122  
 G+QGGSEVTPFLFAIGASLG+++AP LGLFV+LVAALGYT+VFGSATNT PI IG EVFG  
 Sbjct: 319 GYQGGSEVTPFLFAIGASLGIVLAPYLGLFVLVAALGYTTSVFGSATNTFWAFIFIGIEVFG 378

Query: 123 FANTPFYFVIVCLVAYSISHANTYGAQ 149  
 N + + + AY +SH H+TY Q  
 Sbjct: 379 PENALAYPVTSAANVMVSHRHSITYQ 405

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 489

40 A DNA sequence (GBSx0527) was identified in *S.galactiae* <SEQ ID 1563> which encodes the amino acid sequence <SEQ ID 1564>. Analysis of this protein sequence reveals the following:

Possible site: 27  
 >>> Seems to have a cleavable N-term signal seq.

45

| REGION   | LIKELIHOOD | TRANSMEMBRANE | START | END | LENGTH       |
|----------|------------|---------------|-------|-----|--------------|
| INTERCAL | -8.65      | Transmembrane | 47    | 63  | ( 45 - 70)   |
| INTERCAL | -5.04      | Transmembrane | 219   | 235 | ( 208 - 237) |
| INTERCAL | -3.35      | Transmembrane | 168   | 184 | ( 168 - 187) |
| INTERCAL | -0.48      | Transmembrane | 141   | 157 | ( 141 - 157) |

----- Final Results -----

50

| LOCATION            | CERTAINTY            | STATUS  |
|---------------------|----------------------|---------|
| bacterial membrane  | 0.4461 (Affirmative) | < succ> |
| bacterial outside   | 0.0000 (Not Clear)   | < succ> |
| bacterial cytoplasm | 0.0000 (Not Clear)   | < succ> |

55 A related GBS nucleic acid sequence <SEQ ID 9317> which encodes amino acid sequence <SEQ ID 9318> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04382 GB:AP001509 unknown conserved protein in others  
 [Bacillus halodurans]  
 Identities = 75/223 (33%), Positives = 119/223 (52%), Gaps = 18/223 (8%)

-577-

```

5 Query: 17 FSLIIIGGVGALITAVGVKVLLETPSDRYAYLPELVLGVFVIFVYQKGGKGS--- 72
 + +IG +VGA T+ V L R + + +L PL + GL + +Y K G +
 Sbjct: 15 FGIMIGAVIGSATALLLVNDHLGRETNRNPFVL-FLPLAGIALGYLYMKAGTSAGNEL 73

10 Query: 73 YKGVKLVFEVGVGNEETIPKRLVPLVILITWLTFLFGSSAGRGVAVIGATVSHYFQKY 132
 KG LV E G + + R PLV L T + + L FGGS GRG A + Q + G + V +
 Sbjct: 74 YKGVKLVFEVGVGKHEI-IRLGLFVLITFMTPIFGSGTRGGAIVQMGSVAEAVK 132

15 Query: 133 CRINDSQLPIAV-GMAAGFAGLQZPLIATPFAVLVGVLRVIMVPSLIIAALTANF 191
 + + + L+H G+ +NGF F T+ A F E+ +GRG + +P L+A+ +
 Sbjct: 133 FPKYLDITRIILMGGISASGGAAGTPT+TAIPGMNAGLGRGKFAELVPLCLVASPVCHY 192

 Query: 192 VSHSGLGKPFSH-----STATSMALTPIILKLLVLGLCFGL 228
 + + RKF H IAT + + K + + + + F F
 Sbjct: 193 TT-----EKPVHEVEHKFIATVEPSALTSKGVILLATVSL 230

```

There is also homology to SEO ID 1562.

A related GBS gene <SEQ ID 8577> and protein <SEQ ID 8578> were also identified. Analysis of this protein sequence reveals the following:

```

lipop: Possible site: -1 Crend: 5
Mod: Discrim Score: 9.66
GVH: Signal Score (-7.5): -1.12
 Possible site: 27
25 >>> Seems to have a cleavable N-term signal seq.
ALOM program count: 7 value: -10.99 threshold: 0.0
INTEGRAL Likelihood =-10.99 Transmembrane 328 - 344 (314 - 354)
INTEGRAL Likelihood = -8.65 Transmembrane 47 - 63 (-45 - 70)
INTEGRAL Likelihood = -6.32 Transmembrane 255 - 271 (253 - 272)
30 INTEGRAL Likelihood = -4.41 Transmembrane 214 - 230 (208 - 238)
INTEGRAL Likelihood = -3.35 Transmembrane 168 - 184 (168 - 187)
INTEGRAL Likelihood = -2.76 Transmembrane 367 - 383 (367 - 383)
INTEGRAL Likelihood = -0.48 Transmembrane 141 - 157 (141 - 157)
35 PERIPHERAL Likelihood = 0.42 94
modified ALOM score: 2.70

*** Reasoning Step: 3

----- Final Results -----
40 bacterial membrane --- Certainty=0.5394(Affirmative) < success
 bacterial outside --- Certainty=0.0000(Not Clear) < success
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < success

```

The protein has homology with the following sequences in the databases:

```

45 ORF01989[349 - 1491 of 1794]
GP[4512350][db:[BAA75315.1][AB011836(15 - 399 of 424) similar to Bordetella parapertussis
transposase for insertion sequence element(27%-identity)] {Bacillus halodurans}
PIR[T44296][T44296 hypothetical protein [imported] - Bacillus halodurans
%Watch = 15.4
50 %Identity = 33.4 %Similarity = 54.7
Matches = 129 Mismatches = 167 Conservative Sub.s = 82

 222 252 282 312 342 372 402 432
MY*RESKTNINAW*YEQLSKITQNLVPIKKRIIL*TVIKRPIDNIAQQVLVLGGVSGAITAVFGRLVLFITAFRS
 |::||:|:|:|::|
 MNKTPLMLTLTWIFFGIMIGATVGSATAILLTVNDHGLGETFE
 10 20 30 40

 462 492 513 540 570 600 630 660
DYLAILLPFIISIVLGGFLVFVPYQKRG--GKSV-KGMGLVPFYGVGKNKEKITPKRLVPLVILTWTWLTIFGSGAGREGVAQ
 : :: || | : : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
NRFPVPL-PLPLGLGLGYLYMGASAGNELYKGNLNILVESVG-KGMGLMLRGLPVLVLTPTLTIFGSGTGREGQAIIQ
 60 70 80 90 100 110 120

```

[illegible]

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 490

A DNA sequence (GBSx0528) was identified in *S. galactiae* <SEQ ID 1565> which encodes the amino acid sequence <SEQ ID 1566>. Analysis of this protein sequence reveals the following:

```

Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3568(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

40 >GP:AB58234 GB:D67480 chorismate mutase/prephenate dehydratase  
(pheA) [Methanococcus jannaschii]  
Identities = 26/85 (30%), Positives = 46/85 (53%), Gaps = 1/85 (1%)

Query: 2 ELEIEIRQIDEIDQQLVSLSTRNGLIEVAKKHLPLVDNNRNEENLVKKV 61  
14 EIR++IDID+++L+RL+V+++-+DRE++++K+K+  
Sbjct: 4 KLAIRKKIDIDINKLIKLAIRNLSLAKVAVIKNQLGIPIDPEREKYIDRIKCKE 63

45 Query: 62 HQFDVIRATPKDINE--SRVQKE 85  
H D + I + E + + Q +  
Sbjct: 64 HVDVDNIGIKTPOILEIRAAQK 88

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1567> which encodes the amino acid sequence <SEQ ID 1568>. Analysis of this protein sequence reveals the following:

```

Possible site: 37
>>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2356 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

60 An alignment of the GAS and GBS proteins is shown below:

-579-

Identities = 45/91 (49%), Positives = 62/91 (67%)

Query: 1 MELEEIRQELDRIDQQLWSLLETRNGLLIEVIAFKKKHRLPVLNNRRENEVLNNVLKKVQ 60  
 M LE+IRQEI+ ID LV+LLE RM L+ +V A+K + LPVLD REN++L+ V V+  
 Sbjct: 1 MRLEKIRQELNGIDHILVALLERKMAVBSQVATYKLANHLPLVDQARENQILDRVSYLVK 60

Query: 61 NHQFDDVIRATPKDINTESRVYQKENVKSD 91  
 + F+ I TFK IM+ SR YQ +++ GD  
 Sbjct: 61 DQAFEPRIHETFKITMSELRSQVQVQHLLGGD 91

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 491

A DNA sequence (GBSx0529) was identified in *S.galactiae* <SEQ ID 1569> which encodes the amino acid sequence <SEQ ID 1570>. This protein is predicted to be neuraminidase. Analysis of this protein sequence reveals the following:

Possible site: 19  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -3.35 Transmembrane 28 - 44 ( 28 - 47)

----- Final Results -----  
 bacterial membrane --- Certainty=0.2338 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10191> which encodes amino acid sequence <SEQ ID 10192> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA51473 GB:X72967 neuraminidase [Streptococcus pneumoniae]  
 Identities = 294/504 (58%), Positives = 380/504 (75%), Gaps = 10/504 (1%)  
 Query: 303 EDIKSYFOYYCHLNHQLKPKGAILSAKTSVTRGDDGRKKKDNVFGYRI PSLKTKENG 362  
 E+++ Q + + + KLP+GA L+ Kt+++ G G+ NKD + YRIP+LLKt+KGT  
 Sbjct: 299 ESVQKRSQFLKRSIDLEKCLPREGAALTEKTDIFESGRNGKPKHDGKSYRI PALLKTKGT 358  
 Query: 363 LLVAGADERIEQACWQNGVIRRSDDGVTWGKRETI VNLRRNRPVLPVITSGDYSGSPI 422  
 L+ GADER + DWG+IGWIRRSDD+G TWG R TI NLR+NP+ S GSP+  
 Sbjct: 359 LLVAGDERKRLHSSDWDGIGWIRRSDDGKGTWGRVTITNLRRNPKA----SDPSIGSPV 414  
 Query: 423 NDMVALQDTSSTKTRIFSIYIMFPFEGRGVSIANTPEKEYTQIGQSYNLNNGKSK 482  
 N+DM LVQD +TKRIFSIYIMFPFEG+G+ ++ E+ Y +I G++Y LY G+K  
 Sbjct: 415 NIDMVLVQDP--FTRKIFSIYIMFPFEGKIPQMSQKERAYKKIDGTYQILYREGRG- 471  
 Query: 483 VFTIRDKGIVYNPKKCTDYHVTETTESHNLGDIYKQKQLLNIYPTKHKTSPPFLA 542  
 +TIR+ G VY GK TDY V+ + K +S+ GD+YKQ QLLGNIYPT +KTSPPFR+A  
 Sbjct: 472 AYTIRENGFTVYTPDGKATDYRVVVDVVKPAYSDKGDLYKQKQLLNIYPTKHKTSPPFLA 531  
 Query: 543 KSSYVMSYSDDDGRTWSPRDITASLRQGMKFLGIGPGKGI VLVKWPAGRIIIPAYS 602  
 K SY+VMSYSDDDG+TWS+P+DIT ++ MKFLG+GPG GIVL+ GPH GRI+IP Y+  
 Sbjct: 532 KDSYLVMSYSDDDGKTSAPQDITPMVKLVMMKFLGVG+GTGIVLRNGPHKGRILIPYTT 591  
 Query: 603 TNWKSRLHSGSSRLIYSDDHGKVTWTKAVNDRILSNGEKIHSITMDNKKQNTESVD 662  
 TN SHL GSQSR+IYSDDHGKVTW G+AVNDR + +G+KHS TM+N+ QNTES  
 Sbjct: 592 TNNVSHLHSGSSRLIYSDDHGKVTWAGEAVNDRQV-DGQKIHSSTMNRRQAQNTESV 650  
 Query: 663 VQLNGDGIKLFMRNLATGNLAVATSKDGGTWNHVKRYKEVHDVQLSAIRFREDKKEY 722  
 VQL NGD+KLFMR LTH+L+VATSKDGG TW+ +KRY +V D YVQ+SAI H+ KRY  
 Sbjct: 651 VQLNGDVKLFMRGLTGLVATSKDGGVYTWKDIKRYVQVKDVTYVQMSAITHMEGKEY 710

-580-

Query: 723 ILLVNRNGPGKKRQGYARLAQVNRNGSFKMLYHHHIDQGSFAYNSVQQLNRNDKFGVLYE 782  
 I+L NA GP KR++G LA+V NG WL H+ IQ G FAYNS+Q+L N ++G+LYE  
 Sbjct: 711 ILLSNDGSG--KRENGVMVHLARVSENGELTWLAKHNPICQGEFAYNSLQELNGEYIGLYE 768

5 Query: 783 HRRKHGNSFTLVNYKVPNSFLSQN 806  
 H EK QN++TL+++ FNN FLS++  
 Sbjct: 769 HTEGGQNYTLSPKRFNDPLSKD 792

No corresponding DNA sequence was identified in *S.pyogenes*.

- 10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 492

- A DNA sequence (GBSx0530) was identified in *S.galactiae* <SEQ ID 1571> which encodes the amino acid sequence <SEQ ID 1572>. This protein is predicted to be unnamed protein product (gatC). Analysis of this protein sequence reveals the following:

Possible site: 34  
 >>> Seems to have an uncleavable N-term signal seq

|          |                    |               |                        |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood =-12.63 | Transmembrane | 154 - 170 ( 149 - 178) |
| INTEGRAL | Likelihood =-11.99 | Transmembrane | 103 - 119 ( 98 - 123)  |
| INTEGRAL | Likelihood =-7.91  | Transmembrane | 21 - 37 ( 14 - 40)     |
| INTEGRAL | Likelihood =-6.53  | Transmembrane | 448 - 464 ( 444 - 467) |
| INTEGRAL | Likelihood =-5.89  | Transmembrane | 47 - 63 ( 45 - 68)     |
| INTEGRAL | Likelihood =-5.10  | Transmembrane | 356 - 372 ( 352 - 373) |
| INTEGRAL | Likelihood =-4.78  | Transmembrane | 330 - 346 ( 328 - 350) |
| INTEGRAL | Likelihood =-4.41  | Transmembrane | 376 - 392 ( 375 - 393) |
| INTEGRAL | Likelihood =-3.72  | Transmembrane | 243 - 259 ( 235 - 266) |
| INTEGRAL | Likelihood =-2.55  | Transmembrane | 277 - 293 ( 275 - 293) |

----- Final Results -----

|                     |     |                                |         |
|---------------------|-----|--------------------------------|---------|
| bacterial membrane  | --- | Certainty=0.6052 (Affirmative) | < succ> |
| bacterial outside   | --- | Certainty=0.0000 (Not Clear)   | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000 (Not Clear)   | < succ> |

- A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1573> which encodes the amino acid sequence <SEQ ID 1574>. Analysis of this protein sequence reveals the following:

Possible site: 35  
 >>> Seems to have an uncleavable N-term signal seq

|          |                    |               |                        |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood =-12.31 | Transmembrane | 154 - 170 ( 150 - 179) |
| INTEGRAL | Likelihood =-11.68 | Transmembrane | 104 - 120 ( 99 - 124)  |
| INTEGRAL | Likelihood =-9.82  | Transmembrane | 447 - 463 ( 442 - 469) |
| INTEGRAL | Likelihood =-7.91  | Transmembrane | 22 - 38 ( 11 - 41)     |
| INTEGRAL | Likelihood =-7.11  | Transmembrane | 377 - 393 ( 375 - 403) |
| INTEGRAL | Likelihood =-5.89  | Transmembrane | 48 - 64 ( 46 - 69)     |
| INTEGRAL | Likelihood =-4.78  | Transmembrane | 331 - 347 ( 329 - 351) |
| INTEGRAL | Likelihood =-3.88  | Transmembrane | 357 - 373 ( 353 - 373) |
| INTEGRAL | Likelihood =-2.55  | Transmembrane | 278 - 294 ( 276 - 294) |
| INTEGRAL | Likelihood =-1.22  | Transmembrane | 240 - 256 ( 240 - 257) |

----- Final Results -----

|                     |     |                                |         |
|---------------------|-----|--------------------------------|---------|
| bacterial membrane  | --- | Certainty=0.5925 (Affirmative) | < succ> |
| bacterial outside   | --- | Certainty=0.0000 (Not Clear)   | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000 (Not Clear)   | < succ> |

An alignment of the GAS and GBS proteins is shown below:

- Identities = 419/482 (86%), Positives = 447/482 (91%)
- Query: 1 MQVFLNIVNKFPDPIIHMGSGVVMILVMTGLAMIPGVKFSKALSGGIKLAIALTIGIAII 60  
 MQ FL+I+NK I +GSGVVMILVMTGLAMIPGVKFSKALSGGIKLAIALTIGIAII

-581-

Sbjct: 2 MQPFLDIINKILGFPFIQLSSGVMLIVMTGLAMIPGVKPTKALEGGIKLAIALITGIGRII 61

Query: 61 GILTGAFPSLSQAFVKNKGISLIIIDVGNAPLATITWGSPTTYLFLIMLVNIVMIVMK 120  
GILTGAFPSLSQAFVNTGII+L+IIDVGNAPLATITWGSPTTYLFLIMLVNIVMIVMK

5 Sbjct: 62 GILTGAFPSLSQAFVKNKGISLIIIDVGNAPLATITWGSPTTYLFLIMLVNIVMIVMK 121

Query: 121 KTDILDVDIFDIWHLSITGLIIMWYAKGNPLTLLSVIATVAIIFUGVLKIIINSDMKP 180  
KTDILDVDIFDIWHLSITGLIIMWYAKGNPLTLLSVIATVAIIFUGVLKIIINSDMKP

10 Sbjct: 122 KTDILDVDIFDIWHLSITGLIIMWYAKGNPLTLLSVIATVAIIFUGVLKIIINSDMKP 181

Query: 181 TFDLLGTGPTSPWSTHMYNMPIIMVLKLPDKVFPGLDKYDFDAKLNKAGFWGS 240  
TFDLLGTGPTSPWSTHMYNMPIIMVLKLPDKVFPGLDKYDFDAKLNKAGFWGS

15 Sbjct: 182 TFDLLGTGPTSPWSTHMYNMPIIMVLKLPDKVFPGLDKYDFDAKLNKAGFWGS 241

Query: 241 KFFIGMILGLVIGIMNFTFPAALGNWFLGPTAGACLEFLSLIGSWFIAAVERLSQGI 300  
KFFIGM LG VIGIM+P F+ ++ WF LGPTAGACLEFLSLIGSWFIAAVERLSQGI

20 Sbjct: 242 KFFIGMILGLVIGIMNFTFPAALGNWFLGPTAGACLEFLSLIGSWFIAAVERLSQGI 301

Query: 301 TIFNARKHGRRFNIGLWNPFIAGRAEINACNIIAPIMLVEAILLKGKNGILFLAGII 360  
TIFAN+MHGRRFNIGLWNPFIAGRAEINACNIIAPIMLVEAILLKGKNGILFLAGII

25 Sbjct: 302 TIFANARKHGRRFNIGLWNPFIAGRAEINACNIIAPIMLVEAILLKGKNGILFLAGII 361

Query: 361 AMGVTALLVVTIRGLIRMITPTGLLPLFLSGTMIAPFATELAKKVGAFPTAGRAGSL 420  
AMG+TEALLVVTIRGLIRMI PG+ILLPLFLSGTMIAPFATELAKKVGAFPTAGRAGSL

30 Sbjct: 362 AMGVTALLVVTIRGLIRMITPTGLLPLFLSGTMIAPFATELAKKVGAFPTAGRAGSL 421

Query: 421 ITHSTLEGMEKIPGYVIGQATTOQLSIIITLIIIPATYILGLPMWYAKMORRNEYAAN 480  
ITHSTLEGMEKIPGYVIGQATTOQLSIIITLIIIPATYILGLPMWYAKMORRNEYAAN

35 Sbjct: 422 ITHSTLEGMEKIPGYVIGQATTOQLSIIITLIIIPATYILGLPMWYAKMORRNEYAAN 481

Query: 481 QK 482  
K  
Sbjct: 482 MK 483

35 A related GBS gene <SEQ ID 8579> and protein <SEQ ID 8580> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 9  
McG: Discrim Score: 4.31  
GVH: Signal Score (-7.5): -2.64

40 Possible site: 34

>>> Seems to have an uncleavable N-term signal seq

| ALOM program | count:              | 6 value:      | -12.63 threshold:      | 0.0 |
|--------------|---------------------|---------------|------------------------|-----|
| INTEGRAL     | Likelihood = -12.63 | Transmembrane | 154 - 170 ( 149 - 178) |     |
| INTEGRAL     | Likelihood = -11.99 | Transmembrane | 103 - 119 ( 98 - 123)  |     |
| INTEGRAL     | Likelihood = -7.91  | Transmembrane | 21 - 37 ( 14 - 40)     |     |
| INTEGRAL     | Likelihood = -5.89  | Transmembrane | 47 - 63 ( 45 - 68)     |     |
| INTEGRAL     | Likelihood = -4.88  | Transmembrane | 243 - 259 ( 235 - 265) |     |
| INTEGRAL     | Likelihood = -1.22  | Transmembrane | 268 - 284 ( 268 - 284) |     |
| PERIPHERAL   | Likelihood = 0.85   |               | 127                    |     |

50 modified ALOM score: 3.03

\*\*\* Reasoning Step: 3

----- Final Results -----

55 bacterial membrane --- Certainty=0.6052 (Affirmative) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

60 ORF00838 (343 - 1122 of 1455)  
E0AD|91348|RC2092 (9 - 344 of 451) PTS system, galactitol specific IIC component  
[Escherichia coli] OMNI|NT01RC2494 PTS system galactitol-specific enzyme IIC component  
SP|P37189|PTK ECOLI PTS SYSTEM, GALACTITOL-SPECIFIC IIC COMPONENT (IIC-GAT) (GALACTICOL-  
PERMEASE IIC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, C COMPONENT).

65 GP|1736809|dbj|BA015955.1|D90847 PTS system, Galactitol-specific IIC component (IIC-GAT)





-583-

Identities = 85/100 (85%), Positives = 96/100 (96%)

Query: 1 MIKILAACGAGVNSSHQIKDAISTQIGDRGYNVHCDAVMVXDITERMVNXYDITPTPIAKT 60  
 MIKILAACGAGVNSSHQIKDAISTQIGDRGY+VHCDAVMVXDITE+V++YDITPTPIAKT  
 Sbjet: 1 MIKILAACGAGVNSSHQIKDAISTQIGSDRGY/VHCDAVMVXDITEBLVSKYDITPTPIAKT 60

Query: 61 DLGFMVPIPVVEAGPILIRIPVMSSEVPVFTALBQVLIKENIL 100  
 DLGF+PIP+VEAGPILIRIP+MSEVPV LE+VIKE++L  
 Sbjet: 61 DLGFMVPIPVVEAGPILIRIPVMSSEVPFAELSERVIKENIL 100

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 494**

A DNA sequence (GBSx0532) was identified in *S. agalactiae* <SEQ ID 1579> which encodes the amino acid sequence <SEQ ID 1580>. This protein is predicted to be Gata. Analysis of this protein sequence reveals the following:

Possible site: 15  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2078 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10193> which encodes amino acid sequence <SEQ ID 10194> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GF:AA009977 GB:AF248038 Gata (Streptococcus agalactiae)  
 Identities = 39/135 (28%), Positives = 76/135 (55%), Gaps = 9/135 (6%)

Query: 16 QBELFDLVSKALIKQHYVSPNYRQAVKERERFPTGLKIDLDGTPIQVVAIPHITETQYC 75  
 Q L +++S+ L+++ YV + +A+ +RE+++PTGL+++ VAIPHT ++Y  
 Sbjet: 20 QTNLEVLVSQYLLQGVYKTEPSKALLQREKDYPTGLQLE-----NMVAIPHITYSEV 73

Query: 76 LVDRIFYVNSQPITFKEMINPEECKRVQDFPIINSRN-SNQSIDLNLITFTTKENIL 134  
 L I+ K +PI+F M E+E + + ++ N +Q+ +L+ L+T F +  
 Sbjet: 74 LKPFYIYNIGLSPISFIQM-GTEDELVMARYVIVLQISNPKDQAGLLABIMLSPNPKIV 132

Query: 135 DRLNELDNKKERINE 149  
 +L E+ K E + +  
 Sbjet: 133 QQL-EMAQTKKALRN 146

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1581> which encodes the amino acid sequence <SEQ ID 1582>. Analysis of this protein sequence reveals the following:

Possible site: 33  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3130 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 102/154 (66%), Positives = 122/154 (78%)

Query: 4 VTQDILFDIAHSQBELFDLVSKALIKQHYVSPNYRQAVKERERFPTGLKIDLDGTPIQ 63

-584-

```

 V +ILP +A +Q ELFDLV+ L K YV+ Y QA+ ERE FPTGLK<DLKDG+ I
Sbjct: 1 VFNILPTEARTQPELFDLVASHLEKGVFVTQYHQAQVREAVPFTGLKVDLKDGDIL 60

 Query: 64 YVAIPHETSTQYCLVDRIFFYVNSQPTTFKIMINPEBCKVDFFFLINSNSNQSDILSN 123
 Y AIPHETSTYCLVD++ YV+NSQ +TFKIMINPEB+C V DEFFLINS+K Q+ ILSN
Sbjct: 61 YVAIPHETSTKYCLVDQVYVYVNSQALTFKIMINPEBCLVDFFFLINSQNSQSTILSN 120

 Query: 124 LITFPITKGNLDRILHLDGDKGKINHLIKGVF 157
 LITFPITKGNL L L D+K+ I+YLIKGVF
Sbjct: 121 LITFPITKGNLSYLSLKDQKALSNYLIKGVF 154

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 495

- 15 A DNA sequence (GBSx0533) was identified in *S. agalactiae* <SEQ ID 1583> which encodes the amino acid sequence <SEQ ID 1584>. Analysis of this protein sequence reveals the following:

```

Possible site: 61
>>> Seems to have no N-terminal signal sequence

```

- 20 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.1429 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 25 The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAA25176 GB:M60447 repressor protein [Lactococcus lactis]
Identities = 139/255 (54%), Positives = 189/255 (73%), Gaps = 6/255 (2%)

```

- ```

      Query: 1 NLKREHLQKILIEKVNIHIVTVMINSELDVSNITVARNLDELQKAGILLRIH9GAQKN 60
      N K+ RL+KI++ + I+G +T+ EI++ELD+SEMT ERDL L+ GL L R H3GAQ ++
Sbjct: 7 NNKGRLEKILQMLKIDGTTTKEITIDELDISMTARNLDALEADGLLIRTHGAGNLLS 66

      Query: 61 ASPTPQNYEKSNTKYDIQNKLELTAQPAKQFINDGETIFIGRGTTLKELATQLLD--- 117
      + + EK+ EK + T EK+IA+ A I DQ+TIFIGRGTTL +IA +L
Sbjct: 67 SK---KPLEKTHIEKKSIMTKKIDIANKACSLINDGTTIFIGPOTTLVQLALELGRIG 123

      Query: 118 FKIRVITNSLFPVFNILNQSSTLDLILVGGEXREITGAQFVGSVTINSIKSLNPKAFVSN 177
      +KIRV+INSLEVF IIN S T+DL+L+GGEXREITGAQFVGS+ ++K+ P+KAFV +N
Sbjct: 124 YKIRVITNSLFPVFNILNQSSTLDLILVGGEXREITGAQFVGSMASTNLKAMPKAFVIRAN 183

      Query: 178 GVPEKSIATYDEGBEIQRIALNNSPEKFLVDSQKPGKYDFYTFYQLDDIDFVLTDHI 237
      V SIATY + EG IQ+ALNN+ EKFLVDS KP +YDF+ FY LD +D +D+TD+ I
Sbjct: 184 AVTHNSIATYSDKEGVIQQLANNAVEKFLVDSYTKPRYDFPFIINLQDLDTITTENQI 243

      Query: 238 DNVVIGQVYSFTKIL 252
      E++S +T IL
Sbjct: 244 SPQHLESPQYTTIL 256

```

- A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1585> which encodes the amino acid sequence <SEQ ID 1586>. Analysis of this protein sequence reveals the following:

```

Possible site: 60
>>> Seems to have no N-terminal signal sequence

```

- 55 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.0740 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

-585-

Identities = 161/252 (63%), Positives = 195/252 (76%), Gaps = 3/252 (1%)

Query: 1 MLKPERLQKIIERKWNIGIVTVNEIMEDELVSMTVRRLDELKAGKILIRIHGGAQKN 60
 5 Sbjct: 1 MEKRRRL KI E VN GIVTVN+I+ L+VSDMIVRRLDEL+KAG LIRIHGGAQ +
 MLKPERLAKITEIVHQGIVTVNDIIQTIANVSDMIVRRLDELKAGKILIRIHGGAQSIT 60

Query: 61 ASPTPNQYKSNTEKYDIQINRKLKLAQFANQFINOGETIFIGPGTILSKLATQLOLQFKI 120
 10 Sbjct: 61 M---PNKRRSNIEKZVQTEKRWELASYATQLVNDGETIFIGPGTILRCFAQQLANQI 117

Query: 121 RVTNSLFPVNLNQSSSTLDLILVGGSEYREITQAFVGSVTINSKLSNPKAFVSSNGVF 180
 10 Sbjct: 118 R+VTNSLFPVNL S T+DLIL+GGSEY ITGAFVGS+ +I SL F+KAF+S NG++
 RIVTNSLFPVNLQOSETIDLILIGSEYRSTGAFVGSLSQNISSLKPAKAFISQNGIY 177

Query: 181 EKSIIATYDGHGHEIQRIALNNSFEKFLVDSQKFGKYDFTTFYQLDDIDPVLTDMIDNV 240
 15 Sbjct: 178 KNDIATYSETGHEIQLANNSIEKYLVDNQKFNATDFFIFPHLNMIDAVVDSQITD 237

Query: 241 VKEQYSSFTKIL 252
 20 Sbjct: 238 VIERYSFTQLL 249

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 496

A DNA sequence (GBSx0534) was identified in *S. agalactiae* <SEQ ID 1587> which encodes the amino acid sequence <SEQ ID 1588>. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3436 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD13797 GB:AP062533 unknown [Streptococcus agalactiae]

Identities = 86/371 (23%), Positives = 136/371 (36%), Gaps = 79/371 (21%)

40 Query: 11 DLSESELKAAQEFLLGKSEANQDKPKTKTAQELIYEAIEPKAIVKPEDLLFGIAQATDYK 70
 DL++ + L K D TG IEP+ V L AT
 Sbjct: 526 DLTQIAFAEQELMLKDKHYRYDIVDTG-----IEPLAVDVSLLPMAQGNATYDT 576

45 Query: 71 NGTFVIPHDKHYHYVELAWPCEKDLADSCKTYSLEDVLTAKYVMHPEKRPKVEG 130
 +FVIPH DH H V W + +AT KY M HPE RP V W
 Sbjct: 577 GSSFVIPHIDHIVVPYSWLRNQ-----LATIKYVMHPEVRFDV--MS 619

50 Query: 131 KDAEIKKESDNKADKPSAPPTDNKSTNSNDLNSPAEVFKQAKPEKIVPLDKIAAHMA 190
 K + + + P+ P D ++ + SA EV +K + + AA
 Sbjct: 620 KPGH-----EBSGSVIPNVTPIDKRGAMNMQIHSAREV---QKALAGRFPA-- 665

55 Query: 191 YAVGPRDQQLVPHHDIHYHNVPMAMFDEKGLNKAPEGYTLQQLFST--IKYMEHPELE 248
 D I D W D +G +L+ + + + EL
 Sbjct: 666 ----PDGYIPDRDVLAKETFWKDGSPSIFRADGSSSLRTINKSDLSQAEWQQAQELL 719

60 Query: 249 KEKGAGHSDHNKSGSNKNAQVAPORPEKSDSGKVTNHYGFDYVNGKSDSEEP-EKQED 307
 +K G +D +K P+E+ + +K ++ +P E ++
 Sbjct: 720 AKGNAGDATTDPK-----PEKQKQ-----ADKSNENQKPSRASK 754

Query: 308 ESELDIYELGNAQNAKKYGMDDQSPFKQLIQLSNKYSVSFESFNINSGOVVTKIXGSK 367
 E E D++ + YG+DR + E +QL+ K++ + VQ K+G
 Sbjct: 755 EKESDDF-----IDSLPYGLDRATLEHINQIAQKAMID-PKYLIPQREGVQFNKNGEL 809

-586-

Query: 368 VLVDIKTLTEV 378
 V DIKTL ++
 Sbjct: 810 VTYDIKTLQOI 820

A related DNA sequence was identified in *S.agalactiae* <SEQ ID 6983> which encodes the amino acid sequence <SEQ ID 6984>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have an uncleavable N-term signal seq
 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS gene <SEQ ID 8581> and protein <SEQ ID 8582> were also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Crend: 2
 McG: Discrim Score: 6.06
 GWH: Signal Score (-7.5): -5.61
 Possible site: 26
 >>> Seems to have an uncleavable N-term signal seq
 ALCM program count: 0 value: 2.23 threshold: 0.0
 PERIPHERAL Likelihood = 2.23 6
 modified ALCM score: -0.95
 *** Reasoning Step: 3
 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1589> which encodes the amino acid sequence <SEQ ID 1590>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have an uncleavable N-term signal seq
 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 808/825 (97%), Positives = 816/825 (97%), Gaps = 3/825 (0%)

Query: 2 KTTYGYIGSVAAILLATHIGSYQLGKHMMGLATKDNQIAYIDSGKGVKAPTKNTMDQ 60
 Sbjct: 2 KTTYGYIGSVAAILLATHIGSYQLGKHMMG ATKDNQIAYIDSGKGVKAPTKNTMDQ 60

Query: 61 ISAEEGISAEQIVVKITDQGYVTSHGDHYHFNGKVPYDAIISRELLMTDPNYHFKQSDV 120
 Sbjct: 61 ISAEEGISAEQIVVKITDQGYVTSHGDHYHFNGKVPYDAIISRELLMTDPNYHFKQSDV 120

Query: 121 INBILDGYVIKNGNYYVYLKPGSKRKNIRTKQIAEQVAKGTKEAKRKGGLAQVAHLSKE 180
 Sbjct: 121 INBILDGYVIKNGNYYVYLKPGSKRKNIRTKQIAEQVAKGTKEAKRKGGLAQVAHLSKE 180

Query: 181 EVAAVNEAKRQRYYTTDDGYIPSPDTIIDLDGDAYLVPHGNHYHYIPKKDLSPSELAAQ 240
 Sbjct: 181 EVAAVNEAKRQRYYTTDDGYIPSPDTIIDLDGDAYLVPHGNHYHYIPKKDLSPSELAAQ 240

-587-

Query: 241 AYWSQKQGRGARPSDYRPTAP--GRRKAPIPVNTYMPGQGHQPDNGSYHAPPRPNDAS 298
 AYWSQKQGRGARPSDYRPTAP GRRKAPIPVNTYMPGQGHQPDNGSYHAPPRPNDAS
 5 Sbjct: 241 AYWSQKQGRGARPSDYRPTAPAGRRKAPIPVNTYMPGQGHQPDNGSYHAPPRPNDAS 300

Query: 299 QNKHQRDEFKGTFFKSLDQLHRLDLKYNHVEEDGLIPEPTQVTKSNAPGVVPHGDHYH 358
 QNKHQRDEFKGTFFKSLDQLHRLDLKYNHVEEDGLIPEPTQVTKSNAPGVVPHGDHYH
 Sbjct: 301 QNKHQRDEFKGTFFKSLDQLHRLDLKYNHVEEDGLIPEPTQVTKSNAPGVVPHGDHYH 360

Query: 359 IIPRSQLSFLFEMELADRYLAQQTDDNGSGSDHSPKSKFVHTFLGHRIRKAYGKGLDQK 418
 IIPRSQLSFLFEMELADRYLAQQTDDNGSGSDHSPKSKFVHTFLGHRIRKAYGKGLDQK
 10 Sbjct: 361 IIPRSQLSFLFEMELADRYLAQQTDDNGSGSDHSPKSKFVHTFLGHRIRKAYGKGLDQK 420

Query: 419 YDTSDAYVPSKESIHSDKSGVTAKHGMHFRHYTGPELQYELDEVANWVKAKQADELV 478
 YDTSDAYVPSKESIHSDKSGVTAKHGMHFRHYTGPELQYELDEVANWVKAKQADELV
 15 Sbjct: 421 YDTSDAYVPSKESIHSDKSGVTAKHGMHFRHYTGPELQYELDEVANWVKAKQADELV 480

Query: 479 AALDQEGQKEKPLFDTKYVRKVTGDKGVGYMPKDGKDYFYARYQLDLTQIAFAEQELM 538
 AALDQEGQKEKPLFDTKYVRKVTGDKGVGYMPKDGKDYFYARYQLDLTQIAFAEQELM
 20 Sbjct: 481 AALDQEGQKEKPLFDTKYVRKVTGDKGVGYMPKDGKDYFYARYQLDLTQIAFAEQELM 540

Query: 539 LKDKKHRYRDIYDTGIEPLAVLSSLPMEAGNATYDTGSSFVIPHIDHIVVYPSMLTR 598
 LKDKKHRYRDIYDTGIEPLAVLSSLPMEAGNATYDTGSSFVIPHIDHIVVYPSMLTR
 Sbjct: 541 LKDKKHRYRDIYDTGIEPLAVLSSLPMEAGNATYDTGSSFVIPHIDHIVVYPSMLTR 600

Query: 599 NQIATIKYVMQHEVRPDWNSKFGHESSGVIPNVTPLDKRAGMENWQIHSAAEQKAL 658
 +QIATIKYVMQHEVRPDWNSKFGHESSGVIPNVTPLDKRAGMENWQIHSAAEQKAL
 Sbjct: 601 NQIATIKYVMQHEVRPDWNSKFGHESSGVIPNVTPLDKRAGMENWQIHSAAEQKAL 660

Query: 659 AEGRFAPDGYIFDPRDLAKETFWKDGSPSPRADGSSSLRTINKSDLSQAEWQQAQEL 718
 AEGRFAPDGYIFDPRDLAKETFWKDGSPSPRADGSSSLRTINKSDLSQAEWQQAQEL
 30 Sbjct: 661 AEGRFAPDGYIFDPRDLAKETFWKDGSPSPRADGSSSLRTINKSDLSQAEWQQAQEL 720

Query: 719 LAKKAGADATDTDKPKKQADKSNNOQPSEASK EKESSDDFTDSLPHYGLRATLED 777
 LAKKAGADATDTDKPKKQADKSNNOQPSEASK EKESSDDFTDSLPHYGLRATLED
 35 Sbjct: 721 LAKKAGADATDTDKPKKQADKSNNOQPSEASK EKESSDDFTDSLPHYGLRATLED 780

Query: 778 HINQLAQKANIDPKYLIPOPEGVQFYNNKGELVTYDIKTLQQINP 822
 HINQLAQKANIDPKYLIPOPEGVQFYNNKGELVTYDIKTLQQINP
 40 Sbjct: 781 HINQLAQKANIDPKYLIPOPEGVQFYNNKGELVTYDIKTLQQINP 825

SEQ ID 5852 was expressed in *E.coli* in two different forms. GBS293dNterm was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 147 (lane 14; MW 74kDa + lanes 17 & 18; MW 48.8kDa). GBS293C was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figures 148 (lane 2-4; MW 71kDa + lanes 5 & 7; MW 46kDa) and 182 (lane 7; MW 46kDa). Purified GBS293C-His is shown in Figure 241, lanes 8 & 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 497

A DNA sequence (GBSx0535) was identified in *S.galactiae* <SEQ ID 1591> which encodes the amino acid sequence <SEQ ID 1592>. Analysis of this protein sequence reveals the following:

possible site: 23
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

-588-

The protein has homology with the following sequences in the GENPEPT database:

```
>GF:AD13797 GB:AF062533 unknown [Streptococcus agalactiae]
Identical = 213/463 (46%), Positives = 277/463 (59%), Gaps = 41/463 (8%)

5   Query: 4   KKTIV-IISALSVALPGTGVGAYQLGSYNA--QKSDNSVGVVKTDKSDSKAQATAVNKTDP 60
      KKT I +++ L T +G+YQLG ++ DN ++Y+ D S K +A NKT D
      Sbjct: 2   KCTYGYIGSVAAILLATHIGSYQLGHHHGLATKMQIAYI--DDESKVKAPEKTKTMD 59

10  Query: 61   QISKEGISAEBQIVVKITDDGYVPSHGDIHYHYVNGKVPYDAIISSEELIMKDPSPVFNKAD 120
      QIS EEGISAEBQIVVKITD GYVPSHGDIHY+YNGKVPYDAIISSEEL+M DP+Y F ++D
      Sbjct: 60   QISAEBGISAEBQIVVKITDQGYVPSHGDIHYFYNGKVPYDAIISSEELIATDNYHFQKSD 119

      Query: 121  VINEVDGYIIVKNGKYLLKRGSKRTNVRTKEQIQREEWGKGGSGKSGSGKISGART 180
      VINE+ DGY+IKVNG YY+YLG GSKR N+RTK+QI +Q + +K B+ + A+
      Sbjct: 120  VINEILDGYIVKNGYVYLLKPGSKRNIRTKQIARQVAKOTK-----EAKKGLAQV 174

      Query: 181  QALS-----ASVREAKASGRYTTDDGYVPSPTDVIDMGGDAFLVPHGDHPHYIPKADLSPS 236
      LS A+V EAK GRYTDDGY+PSPTD+IDD+GDA+LVPHG+H+HYIPK DLSPS
      Sbjct: 175  AHLSEKEEVAANVEAKRGRTTDDGYVPSPTDIIIDLDGAYLVPHENHYIYPKDLSPS 234

20  Query: 237  ELGAAQAYVNRKTGRSGNS--KPSNBSGYIHASAFSGNVSTGRINAFIPSPRVTHANH 294
      EL+AAQAYV++K GR S +P+ + A P + G+ H
      Sbjct: 235  ELAAQAYVSGKQGRGARPDYRTPAPGRKCAPIDVTPNPGQSHQPD-----NGGYH 288

25  Query: 295  WSKPACNGHATAPKHEAPTTPKINQDSDLNMLKRLYAQPLYARHVESDGLVYDPAQVNAF 354
      + P N A+ KH + K ++L +L+ L RHVE DGL+++P QV
      Sbjct: 289  PAPPFRNDASQNKHQ---RDEPKGKTFKELLDQMLRLLYKRVHEEGDLIFESTQVKE 344

      Query: 355  TAIGVSIPEGNHFPFIHYKDMSPLEL-ATRMVAEHRGHILDALGKIDSTKPKHLSHEP 413
      A G +PHG+H+H I +SPLE+E A R +A G+ D + S
      Sbjct: 345  NAFGYVYVPHGDHYHIIPRQLSPLEMLADRYLA-----GQTDNDGSGDSKSPS 394

      Query: 414  NKE-PHTSESHAVTPKQKQKPK---NQIVYSAGTSEAKK 452
      +KE HT H GKP + V+S + I K
35  Sbjct: 395  DKEVITTFPGHRIKAYGKGLDGKPYDTSDAYVFSKBSIHSDVK 437
```

There is also homology to SEQ ID 1590.

SEQ ID 1592 (GBS94) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 3; MW 52.5kDa).

40 GBS94-His was purified as shown in Figure 194, lane 8.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 498

45 A DNA sequence (GBSx0536) was identified in *S. agalactiae* <SEQ ID 1593> which encodes the amino acid sequence <SEQ ID 1594>. This protein is predicted to be Lmb. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> May be a lipoprotein
```

50 ----- Final Results -----

```
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

55 There is also homology to SEQ IDs 1596 and 5548.

A related GBS gene <SEQ ID 8583> and protein <SEQ ID 8584> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: 22  Crend: 5
MoG: Discrim Score:    13.64
5  GvH: Signal Score (-7.5): -5.75
    Possible site: 24
>>> May be a lipoprotein
ALOM program  count: 0 value:  4.83 threshold:  0.0
    PERIPHERAL Likelihood =  4.83    33
10  modified ALOM score: -1.47

----- Final Results -----
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

SEQ ID 8584 (GBS22) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 14 (lane 6; MW 35kDa).

20 The GBS22-His fusion product was purified (Figure 94A; see also Figure 193, lane 4) and used to immunise mice (lane 2 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 94B), FACS (Figure 94C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoinaccessible on GBS bacteria and that it is an effective protective immunogen.

SEQ ID 8584 (GBS22) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 183 (lane 7 & 8; MW 35kDa).

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 499

A DNA sequence (GBSx0537) was identified in *S.agalactiae* <SEQ ID 1597> which encodes the amino acid sequence <SEQ ID 1598>. Analysis of this protein sequence reveals the following:

```

30  Possible site: 39
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -0.59    Transmembrane    19 - 35 ( 19 - 35)

35  ----- Final Results -----
    bacterial membrane --- Certainty=0.1235(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

40  >GP:CAA51352 GB:X72832 CR1 [Streptococcus equisimilis]
    Identities = 104/145 (71%), Positives = 126/145 (86%)

    Query: 1  MKIIQRVNSQSVSIEDDVGSIERGLVLLVGIAPEDTTEDLAYAVRKITSVRIFSDDEG 60
    MK+++QRV +ASVSI+ + G+I +GL+LLVG+ P+D ED+AYAVRKI +MRIFSD +G
45  Sbjct: 1  MKILVLRCKEASVSIIDGKIAGAINQGLLLLVGVGPDNAEDLAYAVRKIVNRIFSDADG 60

    Query: 61  KKNLSIQDIKGSVLISIQFTLFADTKKGNRPFTGAADPVKANQFYDINORLANHVSVE 120
    KKN SIQDIKGS+LS+SQFTL+ADTKKGNRPFTGAA P A+QFYD FN++LA+ V VR
    Sbjct: 61  KKNQSIQDIKGSILSVSOFTLNADTKKGNRPFTGAAPKPMASOFYDRFNQCLADFPVPE 120

50  Query: 121  TGGFGADMQVSLINDGFTIILVLDTK 145
    G FGADMQVSLINDGFTI+LDTK
    Sbjct: 121  RGVFGADMQVSLINDGFTIILVLDTK 145

```

-590-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1599> which encodes the amino acid sequence <SEQ ID 1600>. Analysis of this protein sequence reveals the following:

```

Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1430 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 103/145 (71%), Positives = 124/145 (85%)

Query: 1 MKIIIRVQASVSIEDVVGSIKGLVLLVGIAPEDTTEDIAVVRKITSMRIFSDDEG 60
      MK+++QEV +ASVSI+ + G+I +GL+LLVG+ P+D ED+AYAVRKI +MRIFSD +G
Sbjct: 1 MKLVLQRVKASVSIDGKLAGAINQLLLLVGVGPDNAEDLAYAVRKIVMRIFSDADG 60

Query: 61 KGNLSIQDKISVLSISQFTLPADTKKGNRPAPTAADPVVQANQPYDIPNQLANRVSVE 120
      KGN SIQDKIS+LS+SQFTL+ADTKKGNRPAPTAAD P A+Q YD FN++LA V VE
Sbjct: 61 KGNISIQDKISLISVSQFTLPADTKKGNRPAPTAADKFLASQLYDSFNEGLAEFVVE 120

Query: 121 TSGFGADMQVSLINDGPVTIILDTK 145
      G FGADMQVSLINDGPVTI+LDTK
Sbjct: 121 RGVFGADMQVSLINDGPVTIILDTK 145

```

SEQ ID 1598 (GBS368) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 64 (lane 4; MW 20kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 70 (lane 4; MW 45kDa).

GBS368-GST was purified as shown in Figure 215, lane 6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 500

A DNA sequence (GBSx0538) was identified in *S.galactiae* <SEQ ID 1601> which encodes the amino acid sequence <SEQ ID 1602>. This protein is predicted to be stringent response-like protein (rel) (relA). Analysis of this protein sequence reveals the following:

```

Possible site: 37
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -0.32 Transmembrane 60 - 76 ( 60 - 76)

----- Final Results -----
bacterial membrane --- Certainty=0.1128 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:CHA51353 GB:X72832 stringent response-like protein
[Streptococcus equisimilis]
Identities = 647/739 (87%), Positives = 696/739 (93%), Gaps = 1/739 (0%)

Query: 1 MYKEINLTGEEVVAITSQYMETDVAFKPALNYATAAHYQARKSGEPYIIHPIQVAGI 60
      M KEINLTGEEVVA+ ++YM+ETD A+VK AL+YATAAH+YQ RKGSGEPYI+HPIQVAGI
Sbjct: 1 MAKEINLTGEEVVALAKYMETDAFVKQLDYATAAHFYQVRKSGEPYIHPIQVAGI 60

```


-591-

Query: 61 LADLHLDAVTVACGFLADVVBEDTITLDRITDFGRDVRDI+DGVGTVLKGVEYKSHESQL 120
 Sbjct: 61 LADLHLDAVTVACGFLADVVBEDTITLDRITDFGRDVRDI+DGVGTVLKGVEYKSHESQL 120

5 Query: 121 AENHRKMLMAMSKDIRVILVKLADRLHNMETLHLAKDKQERISRETMEIYAPLAHRLGI 180
 Sbjct: 121 AENHRKMLMAMSKDIRVILVKLADRLHNMETLHLAKDKQERISRETMEIYAPLAHRLGI 180

10 Query: 181 SRIKWELEDLSFRYLNSTBFYKISHMMSEKRERERELVDIIVKIRSYTERQQLYGDLYG 240
 Sbjct: 181 SRIKWELEDLSFRYLNSTBFYKISHMMSEKRERERELVDIIVKIRSYTERQQLYGDLYG 240

Query: 241 RPKHIYSIYRMRDKKKRFDQIYDLIAIRC+MET SDVYAMVGTYHELWRMPGRFKDYI 300
 Sbjct: 241 RPKHIYSIYRMRDKKKRFDQIYDLIAIRC+MET SDVYAMVGTYHELWRMPGRFKDYI 300

15 Query: 301 AAPKANGYQSIHTTYGPGPIBIQIRTKEMQVAEPQVAHNAWKKGTSKVQAQBSV 360
 Sbjct: 301 AAPKANGYQSIHTTYGPGPIBIQIRTKEMQVAEPQVAHNAWKKGTSKVQAQBSV 360

20 Query: 361 GHWIKELVELQDSK-DAKDFVDSVKEDIPTIRIYVPTNGAVQELPRESGPIDPAYAI 419
 Sbjct: 361 GHWIKELVELQDSK-DAKDFVDSVKEDIPTIRIYVPTNGAVQELPRESGPIDPAYAI 420

25 Query: 420 HTQVGEKATCAKNGRMVPLTAIKLTCGVVEITNPNPSPGRDMIK+VTKNKARNKIRQ 479
 Sbjct: 421 HTQVGEKATCAKNGRMVPLTAIKLTCGVVEITNPNPSPGRDMIK+VTKNKARNKIRQ 480

30 Query: 480 FPKQDKETSINKGRELLVDYPOEGVYVFNKYLKDHIEILPRVSVKSEALYAAVQFG 539
 Sbjct: 481 FPKQDKELSVNKGRLMLVYPOEGVYVFNKYLKDHIEILPRVSVKSEALYAAVQFG 540

Query: 540 DLSPISTFNKITEKEREERAKAAAEALNGEIKTHKRDVLKVSSENGVITQASG 599
 Sbjct: 541 DISPVSVFNKITEKEREERAKAAAEALNGEIKTHKRDVLKVSSENGVITQASG 600

35 Query: 600 LLMRIAKCNPVPGDLIEGYITKRGVAIHRSDQNLKSGENYBQRLIDVEMDDGSKKE 659
 Sbjct: 601 LLMRIAKCNPVPGDLIEGYITKRGVAIHRSDQNLKSGENYBQRLIDVEMDDGSKKE 660

40 Query: 660 YMAEIDIYGLNRGLLNDVLQTLNSATKLVTNVNAOPTKMKFANIHVSPGISNLALQIT 719
 Sbjct: 661 YMAEIDIYGLNRGLLNDVLQTLNSATKLVTNVNAOPTKMKFANIHVSPGISNLALQIT 720

45 Query: 720 VVDKIKIIPDVYSVKRTNG 738
 Sbjct: 721 VVERIKAVFDVYSVKRTNG 739

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1603> which encodes the amino acid sequence <SEQ ID 1604>. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.32 Transmembrane 64 - 80 (64 - 80)

50 ----- Final Results -----
 bacterial membrane --- Certainty=0.1128 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

60 The protein has homology with the following sequences in the databases:
 >GP:CAA51353 GB:X72832 stringent response-like protein
 [Streptococcus equisimilis]
 Identities = 700/739 (94%), Positives = 721/739 (96%)

65 Query: 5 MAKIMNVTGEEVIALATYMTKADVAFVAKALAYATAAHFYQVRKSGEPYIVHPIQVAGI 64

-592-

		MAK +N+TGEEV+ALAA YM + D APV KAL YATAAHFYQVRKSGSEFYIHPHQVAGI
	Sbjct: 1	MAKINILTGEEVVALAAKYNMTDAAPVKKALAYATAAHFYQVRKSGSEFYIHPHQVAGI 60
5	Query: 65	LADLHLDAVTVACGFLHDVVEDTITLDRLEADPGHDARDIDVGVTIKLGEVYKSHSEQL 124
		LADLHLDAVTVACGFLHDVVEDTITLDRLEADPGHDARDIDVGVTIKLGEVYKSHSEQL 120
	Sbjct: 61	LADLHLDAVTVACGFLHDVVEDTITLDRLEADPGHDARDIDVGVTIKLGEVYKSHSEQL 120
10	Query: 125	AENHRKMLAMSKDIRVILVKLADRLNNMRTLKLRLKDKQERISRETMEIYAPLAHLRIGI 184
		AENHRKMLAMSKDIRVILVKLADRLNNMRTLKLRLKDKQERISRETMEIYAPLAHLRIGI 180
	Sbjct: 121	AENHRKMLAMSKDIRVILVKLADRLNNMRTLKLRLKDKQERISRETMEIYAPLAHLRIGI 180
	Query: 185	SRKWELEDLAPRYLNETEYFKISHHMKKKRRERREALVAVLKVYTTTQGLGPDVYG 244
		SRKWELEDLAPRYLNETEYFKISHHMKKKRRERREALVAVLKVYTTTQGLGPDVYG 240
	Sbjct: 181	SRKWELEDLAPRYLNETEYFKISHHMKKKRRERREALVAVLKVYTTTQGLGPDVYG 240
15	Query: 245	RPKHIYSIYRRMRDKKKRFDQIFDLIALRCVMETQSDVYAMGVYIHELNRMPGPKFYI 304
		RPKHIYSIYRRMRDKKKRFDQIFDLIALRCVMETQSDVYAMGVYIHELNRMPGPKFYI 300
	Sbjct: 241	RPKHIYSIYRRMRDKKKRFDQIFDLIALRCVMETQSDVYAMGVYIHELNRMPGPKFYI 300
20	Query: 305	AAPKANGYQSIHTTVYGKPGPIELQIRTKDMHQVAEYGAHAWYKKGVRGKVNQAQV 364
		AAPKANGYQSIHTTVYGKPGPIELQIRTKDMHQVAEYGAHAWYKKGVRGKVNQAQV 360
	Sbjct: 301	AAPKANGYQSIHTTVYGKPGPIELQIRTKDMHQVAEYGAHAWYKKGVRGKVNQAQV 360
25	Query: 365	GQNIKELVELQDASINGDAVFDVSKVEDIFSERIYVFTPTGAQVELPK+SGPIDFAYAI 424
		GQNIKELVELQDASINGDAVFDVSKVEDIFSERIYVFTPTGAQVELPK+SGPIDFAYAI 420
	Sbjct: 361	GQNIKELVELQDASINGDAVFDVSKVEDIFSERIYVFTPTGAQVELPK+SGPIDFAYAI 420
30	Query: 425	HTQIGEKATGAKVNGRMVPLTAKLKTGDVVLEITNANSFGPSRDWVLKVTNKARNKIRQ 484
		HTQIGEKATGAKVNGRMVPLTAKLKTGDVVLEITNANSFGPSRDWVLKVTNKARNKIRQ 480
	Sbjct: 421	HTQIGEKATGAKVNGRMVPLTAKLKTGDVVLEITNANSFGPSRDWVLKVTNKARNKIRQ 480
35	Query: 485	FFKNQDKELSVNKGRLVSYFQGGQVANKYLDKGRIBAILPKVSVKSESLYAAVGF 544
		FFKNQDKELSVNKGRLVSYFQGGQVANKYLDKGRIBAILPKVSVKSESLYAAVGF 540
	Sbjct: 481	FFKNQDKELSVNKGRLVSYFQGGQVANKYLDKGRIBAILPKVSVKSESLYAAVGF 540
40	Query: 545	DISPISVFNKITEKERRREBRAKAKABEELVGGEGVHKNKDLKVRSENGVIQAGSG 604
		DISPISVFNKITEKERRREBRAKAKABEELVGGEGVHKNKDLKVRSENGVIQAGSG 600
	Sbjct: 541	DISPISVFNKITEKERRREBRAKAKABEELVGGEGVHKNKDLKVRSENGVIQAGSG 600
45	Query: 605	LLMRIAACCNFVPGDPIGTYITKRGIAIHRSDCHNIKSDQGYQERLIEVWDLNNSKD 664
		LLMRIAACCNFVPGDPIGTYITKRGIAIHRSDCHNIKSDQGYQERLIEVWDLNNSKD 660
	Sbjct: 601	LLMRIAACCNFVPGDPIGTYITKRGIAIHRSDCHNIKSDQGYQERLIEVWDLNNSKD 660
50	Query: 665	YQAEIDYGLNRGGLNDVLQILSNSTKISISTVNAQPTDKMFANIHVSFGI PNL/HLIT 724
		YQAEIDYGLNR GGLNDVLQILSNSTKISISTVNAQPTDKMFANIHVSFGI PNL/HLIT 720
	Sbjct: 661	YQAEIDYGLNRGGLNDVLQILSNSTKISISTVNAQPTDKMFANIHVSFGI PNL/HLIT 720
	Query: 725	VVEKIKAVPDVYSVKRTNG 743
		VVEKIKAVPDVYSVKRTNG 740
	Sbjct: 721	VVEKIKAVPDVYSVKRTNG 739

An alignment of the GAS and GBS proteins is shown below:

Identities = 635/739 (85%), Positives = 691/739 (92%), Gaps = 1/739 (0%)

55	Query: 1	MVKEINLTGEEVVALTSQYMSIEDVAFVKFALNYATAHYQARKSGSEFYIHPHQVAGI 60
		M K +N+TGEEV+ + YM+ + DVAFP AL YATAAH+YQ RSKSGSEFYI+HPHQVAGI 64
	Sbjct: 5	MAKINNTGEEVVALATYMTKADVAFVKALAYATAAHFYQVRKSGSEFYIHPHQVAGI 64
60	Query: 61	LADLHLDAVTVACGFLHDVVEDTITLDRLEADPGHDARDIDVGVTIKLGEVYKSHSEQL 120
		LADLHLDAVTVACGFLHDVVEDTITLDRLEADPGHDARDIDVGVTIKLGEVYKSHSEQL 120
	Sbjct: 65	LADLHLDAVTVACGFLHDVVEDTITLDRLEADPGHDARDIDVGVTIKLGEVYKSHSEQL 124
65	Query: 121	AENHRKMLAMSKDIRVILVKLADRLNNMRTLKLRLKDKQERISRETMEIYAPLAHLRIGI 180
		AENHRKMLAMSKDIRVILVKLADRLNNMRTLKLRLKDKQERISRETMEIYAPLAHLRIGI 180
	Sbjct: 125	AENHRKMLAMSKDIRVILVKLADRLNNMRTLKLRLKDKQERISRETMEIYAPLAHLRIGI 184
	Query: 181	SRKWELEDLAPRYLNETEYFKISHHMKKKRRERREALVAVLKVYTTTQGLGPDVYG 240

-593-

SRIKWELEDL+PRYIAETEPYKISHMM EKRRERE LV+ IV K++YT +QGL+GD+YG
 Sbjct: 165 SRIKWELEDLAPRYIAETEPYKISHMMKKRREREALVEALVSKVITTTQQLGPGDVG 244
 5 Query: 241 RPKHIYSIYKMRDKKKRFDQIYDLIAIRCIMETASDVMVGVYIHELMRPMGRFKDYI 300
 RPKHIYSIYKMRDKKKRFDQI+DLIAIRC+MET SDVMVGVYIHELMRPMGRFKDYI
 Sbjct: 245 RPKHIYSIYKMRDKKKRFDQIYDLIAIRCVMETQSDVMVGVYIHELMRPMGRFKDYI 304
 Query: 301 AAPKANGYQSIHTTVYVGPKEPIEQIRTKEMHQAEPGVAHWAYKGGITSKVNQAQGV 360
 AAPKANGYQSIHTTVYVGPKEPIEQIRTK+MHQVAB+GVAHWAYKGG+ KVNQAQGV
 Sbjct: 305 AAPKANGYQSIHTTVYVGPKEPIEQIRTKDMHQAEPGVAHWAYKGGITSGKVNQAQGV 364
 10 Query: 361 GQWIKQELVELQDESK-DAKDPVDSVKEDIPTERIYVFTPGAVQZLPRESGPIDFAYAI 419
 GM WI+ELVELQD S DA DPVDSVKEDI+ERIYVFTPGAVQZLP+ESGPIDFAYAI
 Sbjct: 365 GQWIKQELVELQDASNGDAVDVDSVKEDIPTERIYVFTPGAVQZLPKESGPIDFAYAI 424
 15 Query: 420 HTQIGKATGAKVNGRMVPLTAKLKTGDVVEIITNPSFGPSRDWIKLVKTNKARNKIRQ 479
 HTQIGKATGAKVNGRMVPLTAKLKTGDVVEIITN NSFGPSRDW+K+KLVKTNKARNKIRQ
 Sbjct: 425 HTQIGKATGAKVNGRMVPLTAKLKTGDVVEIITNNSFGPSRDWIKLVKTNKARNKIRQ 484
 20 Query: 480 FFFNQDKETSINKGRELLVDYFQBGQYVFNKYLKDKHIEILPRVSVSESLYAAVGF 539
 FFFNQDKES+K+NGKR+LLV YFQBGQYVFNKYLKDKHIEILPRVSVSESLYAAVGF
 Sbjct: 485 FFFNQDKELSVNKGRLDLYSVFQBGQYVFNKYLKDKHIEILPRVSVSESLYAAVGF 544
 25 Query: 540 DLSPTISPNKLTKEKRRERERAKAKAEADKLINGEIKTKRDVLKVSNGVITQASG 599
 D+SPIS+PNKLTKEKRRERERAKAKAE+EL+ GGE+K + +DVLK+V+SENGVITQASG
 Sbjct: 545 DISPTISPNKLTKEKRRERERAKAKAEAEELVKGGEVGHENKVLKVSNGVITQASG 604
 Query: 600 LLMRIACCNPPVPGDLIBGYITKRGVAIHRSDCQNLKSQENYBQRLIDVEMDDSGKSE 659
 LLMRIACCNPPVPGD I+GYITKRG+AIHRSDC N+KSQ+ Y++RLI+VEMD D S K+
 30 Sbjct: 605 LLMRIACCNPPVPGDPIIDGYITKRGIAIHRSDCNHSQDGYOERLIEVEMDDNSKSD 664
 Query: 660 YMABIDIYGLNRSGLLNDVLQTLNATKLSTVNAQPTKDMKFANIHVSPIGLNLAQLIT 719
 Y ABIDIYGLNRSGLLNDVLQ LSN+TK +STVNAQPTKDMKFANIHVSPIGL NL LIT
 Sbjct: 665 YQABIDIYGLNRSGLLNDVLQTLNATKLSTVNAQPTKDMKFANIHVSPIGLNLHLIT 724
 35 Query: 720 VVDKIKIIPDVYSVKRTNG 738
 VV+KIK +PDVYSVKRTNG
 Sbjct: 725 VVEKIKAVEDVYSVKRTNG 743

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 501

A DNA sequence (GBSx0539) was identified in *S.agalactiae* <SEQ ID 1605> which encodes the amino acid sequence <SEQ ID 1606>. This protein is predicted to be 2',3'-cyclic-nucleotide 2'-phosphodiesterase precursor (cpdB). Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -5.79 Transmembrane 779 - 795 (778 - 797)
 50 ----- Final Results -----
 bacterial membrane --- Certainty=0.3314 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

55 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12613 GB:299108 similar to 2',3'-cyclic-nucleotide
 2'-phosphodiesterase [Bacillus subtilis]
 Identities = 297/630 (47%), Positives = 419/630 (66%), Gaps = 21/630 (3%)
 60 Query: 102 KVDLRIMSTTDLHTNLVNDYDYQKESQIGLAKTAVLIREAKENPNTVLVDNGDVIQ 161
 +V L I++TDD+H N+++YDYY DKE+ GLA+TA LI++ ++NPNT+LVNDGSD+IQG

-594-

- Sbjct: 42 QVELSILATIDIHANMDYDYSDKSTADPLARTAQLIQKREQNPNTLLAVDNEDLIQG 101
- Query: 162 TELGTYKAIKVE---VARNSEHPMYQAMNALGYDASTLGNHEFPYGLDYLKIIATANLP 218
ELG Y + ++ + HP+ MMAL YDA TLGHEFPYGLD+L I A+ P
- 5 Sbjct: 102 NPLGEXAVKYQXDDIISGKTTHPIISVMNALKYDAGTLGNHEFPYGLDLOGTIRGADFP 161
- Query: 219 ILNANVLDPKTHOPVFKTYDIIITKPKDSTRGRAVALNIGITGIIVPQILANWKNLEBKV 278
I+ANV + + + Y I K T D G + +G G VPPQI+ WDK NLEB+V
- 10 Sbjct: 162 IVNANVET-TSGENRYTPPVINEKTLIDENGNEQKVKGYGIVFPVQIMWDKKLEBQV 220
- Query: 279 IVKDSVKAIEIIVPMRAKGADVILVLSHSGIDDRYERBGEKNVGYQIAS- IKGVDVAVT 337
V+D + + + E + P M+A+GADV+ L+H+GI Q EN + +A+ KG+DA+++
- Sbjct: 221 QVQDIVESANETIPKMKABGADVIALAFTGIEKQAGSGAENAVFLATKTEGIDAIL 280
- 15 Query: 338 GHSHAFPSGNTGTGYEYKTGVGDGIN--GKINGTPVPMAGKYGDHLGILDLGLSYTNGK 394
GH H FPS +Y GV N G ING PV M +G +LG+IDL L +G
- Sbjct: 281 GHQHLGPFSA-----EYAGVAQFNVEKTINGIIPVMPSSMGKYLGVIDLKLEKADGS 333
- 20 Query: 395 WQVSESSAKIRKIDMNSTTADERIILALAKEAHGDTINVVEQVGTTPAPITSYFALVKDD 454
W+V+S I I N T+ +E + ++ H T+ YVR+ VG T A I S+FA VKDD
- Sbjct: 334 WKVADSKSIESIAGNVTSRNEIVTNTIQTHQMLEYVRKPGKTEADINSFFAQVKDD 393
- Query: 455 PSVOIVNNAQWYVANELKGTPEANPLLSAAPPKAGTSGDATAYTDIPAGPVAIKVA 514
PS+QIV +AQ+WY E+K T NLP+LSA APPKAG R A YT+IPAG +AIONV
- 25 Sbjct: 394 PSQIVTDAQKWYAKENKLTETXNPLISAGAPFAGGRGNAYTYNIPAGDLAIKNGV 453
- Query: 515 DLVLYNDVALLKVTGADLRWLEMSAGQFNQIDPNKAPQNLINTEYTYNFDVLDGLT 574
DLVLYDN ++K+TG++++WLEMSAGQFNQIDP Q ++N +R+YNFDVING+T
- 30 Sbjct: 454 DLVLYNVTQVIVKLGSEVKWLEMSAGQFNQIDPAKSGDQALLNPNRSYNFDVLDGVT 513
- Query: 575 YKFDITOPKNKNDKGVVMSQASQVRDLWNGKFPVADKQRFMLVINNYRASGTFPFGAGIA 634
Y+ D+T+P KVN++GK+M+ +SR+ +L Y GKP++ QEF++VTNRYRASG G+P
- 35 Sbjct: 514 YQVDVTKPAKTNENKVTNADSSRIINLSYEGKPISSQEFVLVTNRYRASGG-GOPHIL 572
- Query: 635 TMNRLN---LENRTQIINYIISEKTINPTADNMGPTESIKDLRLPOTAKAQNLTN 691
T + + + + + +ENRQ + + + YII + +T+NP ADNNW + +L F+ + + AK
- 40 Sbjct: 573 TSDKIVHSGAVENRQVLMQYIIIEQKVTNPDADNMSIA-PVSGTNLTPESILLAKFPADK 631
- Query: 692 SKDIQYIASSTKDESGDYRFVYTEQEKVD 721
+ D+ Y+ S +EG+G Y+ + + D
- 45 Sbjct: 632 ADDVAYVGKSA-NBSYGVYKIQFDDSDNP 660
Identities = 133/567 (23%), Positives = 214/567 (37%), Gaps = 147/567 (25%)
- Query: 104 DLRLMSTDLHLNLYNDYQDKESQKIGLAKTAVLIESAKKENPNTLVNDGDIVIGTTP 163
DL +N T D H + L + A+ I E + E + +L+D GDV G
- 50 Sbjct: 668 DLTVMHTMDTHAHLDD-----AARRMTKINEVRSTNRHLLDAGDVGSGD- 713
- Query: 164 LGTYKAIKPVARNSEHPMYQAMNALGYDASTLGNHEFPYGLDYLKIIATANLP 216
Y +A+ + MN +GVDA T QNHEP+ G D+L AT+
- 55 Sbjct: 714 --LYFTKNGILAD-----LWMNNMGVDAWTFGNHEDKPTVLSDPLSGNSATVDNPN 765
- Query: 217 -----LEILNANVLDPKTHOPVFKTYDIIITKTF-----KDSTRGRAVALNIGITG-IV 262
PI++ANV +++P K++ +TF K G + + + G
- 60 Sbjct: 766 RYHFEAPEPEFVYANV--DVSNEPKLSFVKKPQTFPAGEKKRAGIHPTILLDVGKVA 823
- Query: 263 PFQILWMDKANLE--GKIV-----KDSVKAIEIIVPMRAKGADVILVLSHSGIDG 312
+ D A GK IV ++VKAI+E + + I+ L+H G
- 65 Sbjct: 824 VEGHITTEDTATTSPPKSVFENDAPETKQNTVKAIQ-----EEKVNKIALTHIG-- 874
- Query: 313 DRYBGEENVGQIA-SIKGVDAVYGHSHAEFPSSGNTGTGYEYKTGVGDGINKINTP- 370
N ++A+ +KG+D ++ GH+H T VD + N P
- 70 Sbjct: 875 -----HNRDLAKKVVGDILITGHHH-----TLVDKMEVWNEEPT 912
- Query: 371 -VTMAGKYGDHLGILDLGLSYTNGKQVSESSAKIRKIDMNSTTADERIILALAKEAHGDT 429
V A +YG LG +D+ G Q +S+ + ID ++ E AK+ D
- 75 Sbjct: 913 IVAQAKYEQGLGRVDAVD-EGVVGQTKNSLSVLPIDEHTENPE-----AQLELDQ 966
- Query: 430 INYV---RQOVGTTPAPITSYFALVKDDPSQIVNNAQWYVANELKGTPEANPLLSA 465

-595-

```

      N +      ++VG T      + + QR +V +      + A
Sbjct: 967 KKELEEDVKEKVGYT-----DVALDQQRSHVTRKTNLGNFIADGMLA 1009

5  Query: 486 AAPFKATGEGDAT---AYTIDIPAGFVAIKNVADLYLDNVIALKVTGADLREHLEMSA 541
      A AG R T      I G + + V ++ + N + + TG ++E LE
Sbjct: 1010 KAKEAAGARIAITNGGGIRAGIDKGDITLGEVLNMPFGNTLYVADITGKQIKERLE--- 1066

10 Query: 542 GQFNQIDFNNKAPQILDFEYRTYNFDVIGLTYKFDITQPNKYKIDGKVVNSQASRVRD 601
      Q + N E      F + G + Y F +      NK G      + V+
Sbjct: 1067 -----QGLSNVENGCGAPFQVAGIETFTYIN----NKPG---HRVLEVKI 1104

Query: 602 LMYNGKIFVADKQE--FMIVIMNYRASG 626
      NG VA + + + TNN+ +G
Sbjct: 1105 ESPNGDKVAINTDDTYRVAINFVCGAG 1131

```

There is also homology to SEQ ID 1608. A related sequence was also identified in GAS <SEQ ID 9129> which encodes the amino acid sequence <SEQ ID 9130>. Analysis of this protein sequence reveals the following:

```

20  Possible cleavage site: 27

    >>> Seems to have an uncleavable N-term signal seq
        INTEGRAL    Likelihood = -4.67    Transmembrane    649 - 665 ( 648 - 666)
        INTEGRAL    Likelihood = -2.02    Transmembrane    6 - 22 ( 5 - 22)
        PERIPHERAL  Likelihood = 1.85

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8585> and protein <SEQ ID 8586> were also identified. Analysis of this protein sequence reveals the following:

```

30  Lipop: Possible site: -1    Crend: 7
    MoG: Discrim Score:      6.68
    GVH: Signal Score (-7.5): 0.87
        Possible site: 28
    >>> Seems to have a cleavable N-term signal seq.
35  ALOM program count: 1 value: -5.79 threshold: 0.0
        INTEGRAL    Likelihood = -5.79    Transmembrane    779 - 795 ( 778 - 797)
        PERIPHERAL  Likelihood = 0.53      251
        modified ALOM score: 1.66

40  *** Reasoning Step: 3

    ----- Final Results -----
        bacterial membrane --- Certainty=0.3314(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
45  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

    LEXTG motif: 769-773

```

The protein has homology with the following sequences in the databases:

```

50  ORF01378(298 - 2337 of 3000)
    GP[6782402[emb|CAF70615.1|AJ133440(1 - 680 of 683) cyclo-nucleotide phosphodiesterase,
    putative [Strept
    ooccus dysgalactiae subsp. equisimilis]
    %Match = 38.3
55  %Identity = 59.0 %Similarity = 72.3
    Matches = 403 Mismatches = 181 Conservative Sub.s = 91

    105      135      165      195      225      255      285      315
    LFVHFLT*K*KKLEAQKELATY*NCLTKLSFINGLPLV*SLKTIIRK*D*LVVPNKL**PL*DNIIHVNF*WRRFMSKHY
    60                                     |:|
                                           MDTGY

```


-597-

```

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0296(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 503

A DNA sequence (GBSx0541) was identified in *S.agalactiae* <SEQ ID 1611> which encodes the amino acid sequence <SEQ ID 1612>. Analysis of this protein sequence reveals the following:

```

Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1504(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10195> which encodes amino acid sequence <SEQ ID 10196> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

>GF:CAB12860 GB:Z99109 similar to glucanase [Bacillus subtilis]
Identities = 212/345 (61%), Positives = 268/345 (77%), Gaps = 1/345 (0%)

Query: 30 SMETTLAYIKTLTIPSIPSGTGTQIMTYIIKELRAPGYSPIRTNKGGVMVSLGKNDTKH 89
S+ T+ IK L SIPSPTG T ++ YI L+ + +R +KGG++ +L G++ ++H
Sbjct: 3 SVRKTMELIKELVSIPTGNTYEVINIESLLKEMKVETVRNHHKGLIATLFGRLTSH 62

Query: 90 RMITAHLDTLGMVRAIKPDRKIDKIDGTYTNAIDGNCTIHLGKQKEISGOTALIHQ 149
RM+TAH+DTLGMV+ IK DGRKIDKIDG+ YH+IEGE C I + +GK +GT L+HQ
Sbjct: 63 RMITAHLDTLGMVKEIKDGRKIDKIDGGRYNSIESEYQIETA-SGMVYGTILMHQ 121

Query: 150 TSVHVVYKDAKTAERWQNMIEIRLDEKVTDAETRALGIQVDFISFDPTIITDSGFIKS 209
TSVHVVYKDAK AERWQ NMIEIRLDE V +T LGI VQDF+SFDR IT SGFIKS
Sbjct: 122 TSVHVVYKDAKTAERWQNMIEIRLDEPVCRCRQTEHLGIGVQDFVSFDPRVEITSSGFIKS 181

Query: 210 RYLLDKVSGIIMELLGVYKKEDIQLPYTHIFPSAPRELGHGANGSIPNETVEYLAVDM 269
R+LDDK S +L+ L+ + RDI+LPYTHF S RE+G+G NS+IP ETVEYLAVDM
Sbjct: 182 RHLDDKASVALLRLIHEIQTEIDRLPYTHIFLISNNEBIGGNGNIPPETVEYLAVDM 241

Query: 270 GAKGIDQETDEYTVSICVKDASGPYHYELRCHLAVLAEENIPYKLDIYPPYGSASASNA 329
GA+GD Q TOKY+VEICVKDASGPYHY+LRHLV LAE ++I YKLDIYPPYGSASASNA+
Sbjct: 242 GAIGDQATDEYTVSICVKDASGPYHYQLRKHLVQLAKHHIDYKLDIYPPYGSASASAI 301

Query: 330 RACARVKHALLGAGIESSHGYSRTHIDSIQNTLLVDAYLKENMV 374
++G ++ H L+G GI++SH++ERTH S++ T L+ Y++S MV
Sbjct: 302 KSGHDIVHGLIGPQIDASHAFERTHKSSLRFTAKLLXYVYQSPMW 346

```

There is also homology to SEQ ID 424.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 504

A DNA sequence (GBSx0542) was identified in *S. agalactiae* <SEQ ID 1613> which encodes the amino acid sequence <SEQ ID 1614>. Analysis of this protein sequence reveals the following:

```

Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3157 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAF11472 GB:AE002031 conserved hypothetical protein [Deinococcus radiodurans]
Identities = 55/150 (36%), Positives = 85/150 (56%), Gaps = 2/150 (1%)

Query: 5 LIIIRGNISASGKSTIAKQLQELSPNTILLSCDYLRREMLSTKDGENTITIFLLINLIN 64
LI++RGNIS SGKS++A+ L+ G + QDYLR +L D I L+ + Y
Sbjct: 23 LIVLRGNISGSKSSVARALREKFGYGLANVEQDYLRREVLLRREHDVAGGNIGLITNRY 82

Query: 65 GYHNCYSIIIEGILRSOWYITPVWKGHLKHNPNNTYAYYYDLSPQETVKRHSRLKSLFEG 124
S + LSGIL S Y P+ + + H + +Y+DL F+ETV+RH+TR ++ +FG
Sbjct: 83 CLSAGSVTVLSEGLFSRHYGPNLERL--HADFGHWYFPLPFESTVRRRHATRPQADFG 140

Query: 125 EDSLAHWLEKDFLKEIPEKILTKAMSLED 154
+ W+ +D L + E+++ A SL D
Sbjct: 141 VQDMQAWFQARDVLPFPVQEQELIGPASSLAD 170

```

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 505

A DNA sequence (GBSx0543) was identified in *S. agalactiae* <SEQ ID 1615> which encodes the amino acid sequence <SEQ ID 1616>. This protein is predicted to be periplasmic-iron-binding protein BirC. Analysis of this protein sequence reveals the following:

```

Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -11.46 Transmembrane 9 - 25 ( 5 - 30)

----- Final Results -----
bacterial membrane --- Certainty=0.5585 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAD18094 GB:U75349 periplasmic-iron-binding protein BirA
[Brachyspira hyodysenteriae] (ver 2)
Identities = 114/331 (34%), Positives = 184/331 (55%), Gaps = 3/331 (0%)

Query: 11 YILLVVSIIPIFVPTYSISQPSKLLPPKELVILSPNSQAILTQTIAPFEKY-GIKVLI 69
+I+ + + + +F S SK LVI + + + + F+ K I V+ +
Sbjct: 4 FIIFCMILSMTLTPYSCSSGDSK--NANSLVICYSHPLDMLNITLDDFKARNPDINVEV 61

Query: 70 QQGTQQLIDRLSKBKGQLKADIFFGNGNYTFESHKALFBSYVSKNVHVIIPDYIHSDTA 129
GTG+L+ R+ E D+ +GG +S LFR+Y S N + + + +
Sbjct: 62 TAGTGELLKRVFAKRNMPGLDVLWGQTLNSVSKSKTDLFFNYTSTNEANIIDEFNTGPGP 121

```